# Pre-Analysis Plan v1.7

# Pre-Analysis Plan

## Collaborative Contact in Tourism: A Randomized Field Experiment in Beach Volleyball

**AEA RCT Registry ID:** [to be assigned] **Version:** 1.7 **Date:** 2026-02-22 **Authors:** Jorge Araña Padilla (ULPGC) [+ co-PI(s)]

**This document is intended to be locked upon submission to the AEA RCT Registry. Any post-registration modifications will be tracked with justifications.**

---

## 1. Introduction

### 1.1 Research Question

Does randomly assigning tourists and local residents to collaborative mixed teams in a beach volleyball tournament increase prosocial cross-group behavior, relative to assignment to homogeneous (same-group) teams?

### 1.2 Hypotheses

**H1 (Collaborative contact).** Participants assigned to mixed teams (MixedTeam = 1) will exhibit higher cross-group prosocial behavior on the omnibus index than participants assigned to homogeneous teams (MixedTeam = 0). Formally: $\beta > 0$ in Equation (1).

**H2 (Generalization — exploratory).** The positive effect of mixed-team assignment will extend to outgroup members who were not the participant's teammate — i.e., collaborative contact generalizes beyond direct interaction partners. Formally: the dyadic specification (Equation 4) will show positive effects on non-teammates from the outgroup.

**H3 (Adversarial contact).** A higher proportion of outgroup opponents (adversarial contact) will have a null or negative effect on cross-group behavior. Formally: $\gamma \leq 0$ in Equation (1).

**Decision rules.** We will consider Hypothesis 1 supported if the coefficient on *MixedTeam* in Equation 1 is positive and statistically significant at the 0.05 level (two-sided), and disconfirmed if the coefficient is non-positive or non-significant. We will consider Hypothesis 1 *precisely null* — informative evidence of no effect — if the 95% confidence interval excludes effects larger than 0.15 standard deviations (the approximate minimum detectable effect) while including zero. For Hypothesis 3, $\gamma \leq 0$ is supported if the coefficient on the proportion of outgroup opponents is non-positive, and disconfirmed if it is positive and significant. Hypothesis 2, being exploratory, does not carry a formal decision rule; we interpret the dyadic specification (Equation 4) as suggestive evidence. In all cases, the omnibus index is the primary inference target; individual component effects are reported for transparency but do not override the omnibus conclusion.

### 1.3 Contribution

This is the first randomized field experiment testing Allport's (1954) contact hypothesis in a tourism context with primarily incentivized behavioral outcome measures. The design replicates Lowe (2021, AER), who demonstrated that collaborative contact increased cross-caste prosocial behavior (+0.23σ) while adversarial contact decreased it (−0.38σ) in cricket leagues in rural India. We transpose this framework to tourist–resident contact in Gran Canaria, providing the causal identification missing from prior observational tourism-contact research (e.g., Purper & Araña, in press; Woosnam, Norman, & Ying, 2009).

---

## 2. Study Design

### 2.1 Setting

Beach volleyball tournaments on Gran Canaria (Canary Islands, Spain), across 4 geographically separated tourism zones: Playa de Las Canteras (Las Palmas), Playa del Inglés (Maspalomas), Playa de Bahía Feliz (San Agustín), and Anfi Beach (Mogán). All tournaments on Gran Canaria only. Tournaments are organized in collaboration with the Federación Canaria de Voleibol and local beach volleyball clubs at each location, providing tournament infrastructure, professional equipment, trained referees, and scheduling support.

### 2.2 Intervention

Participants are recruited for a tournament described as a "study of sports dynamics and cooperation." No mention of prejudice, intergroup relations, or the contact hypothesis. Teams of 2 (standard beach volleyball) play ~8 matches over 3–3.5 weeks.

**Treatment:** Random assignment to mixed teams (1 tourist + 1 resident) vs. homogeneous teams (2 tourists or 2 residents).

**Cross-randomized factor:** Team Pay (€20 bonus per match won, shared) vs. Individual Pay (€10 bonus per match with individual performance above threshold).

**Incentive structure.** Experimental payoffs (tournament prizes, trust game endowments) are paid in cash (euros). Separately, all participants receive identical non-monetary participation gifts (local Canary Islands products and branded beach volleyball apparel from partnering clubs) at baseline and upon Endline-2 completion. These gifts are not an experimental variable.

### 2.3 Allport's Conditions

The design satisfies Allport's (1954) four conditions for optimal intergroup contact:

| Condition | How Satisfied |
|---|---|
| Equal status | Both tourists and residents participate as equal teammates |
| Common goals | Win matches together |
| Intergroup cooperation | 2v2 format requires continuous cooperation |
| Institutional support | Official tournaments organized under the Federación Canaria de Voleibol with professional referees, equipment, and club infrastructure at each location |

**2.4 Timeline per Tournament (~60 Days)**

| Days | Phase | Activity |
|---|---|---|
| 1–6 | Recruitment | Beach, hotel, and club recruitment; screen for ≥21-day stay from tournament start (tourists) |
| 7–12 | Baseline | Volleyball ability testing + baseline survey + social network roster |
| 13–14 | Randomization | Computer-based team assignment and match schedule generation |
| 15 | Announcement | Team assignment and match calendar communicated |
| 16–40 | Tournament | ~8 matches per team (2–3 per week) |
| 35–44 | Endline-1 | Trading task (in person, 10-day trading window; items distributed day 35, overlapping last 6 tournament days + 4 days post-tournament) |
| 45–47 | Buffer | Trading data processing and online survey preparation |
| 48–60 | Endline-2 | Social interaction, team formation, voting, trust game (online for all participants) |

## 3. Sample

### 3.1 Target Population

- **Tourists:** Foreign visitors to Gran Canaria with ≥21 days remaining on the island from the start of tournament play (effectively ~30–33 days at first contact). Predominantly Northern European long-stay visitors during the winter season (October–April).
- **Residents:** Local Canarian residents aged ≥18. Recruited from recreational settings (beaches, university, community boards) rather than competitive volleyball clubs, to minimize skill imbalance with tourists.

### 3.2 Planned Sample Size

- **Base design:** 800 participants (4 locations × 2 temporal waves = 8 tournaments × ~100 recruits each)
- **Scaling option:** 1,200 participants (4 locations × 3 waves = 12 tournaments) if funding allows
- **Per tournament:** ~70 active players (35 teams of 2) + ~30 backups
- **Target balance:** ~50% tourists, ~50% residents per tournament

### 3.3 Inclusion Criteria

1. Age ≥18 years
2. Tourist stay ≥21 days remaining from the start of tournament play (tourists only). **Recruitment fallback:** if fewer than 40 eligible tourist participants are recruited at any location-wave after the planned recruitment period, the threshold is relaxed to ≥18 days remaining from the start of tournament play. Participants recruited under the relaxed criterion are flagged for sensitivity analysis (Section 12.4).
3. Able to participate in recreational beach volleyball (self-assessed)
4. Written informed consent (available in Spanish, English, German, Swedish, or Norwegian; Finnish speakers offered the English version)
5. No prior participation in any wave of this study

### 3.4 Exclusion Criteria

1. Tourist stay <21 days from tournament start (or <18 days if the recruitment fallback is triggered)
2. Knows >3 outgroup participants at baseline (preserves "stranger" condition)
3. Prior participation in any wave at any location
4. Reports knowledge of a prior volleyball study participant on the island (spillover screen)

### 3.5 External Validity Scope

Results generalize to **long-stay tourists who self-select into recreational sport activities** and **residents who voluntarily participate**, not to the broader tourist or resident population. The ≥21-day restriction excludes the majority of tourists (median stay ~7 days). Self-selection into a volleyball tournament implies above-average sociability and openness. These limitations are shared with Lowe (2021), who recruited voluntary cricket players. We will report the recruited sample's demographic profile relative to island-wide tourism statistics (ISTAC/Cabildo de Gran Canaria data).

## 4. Randomization

### 4.1 Procedure

Randomization is computer-based (Stata 18, `randomize` command), conducted in the research office before the tournament begins. For each tournament, a unique random seed is generated from atmospheric noise (random.org) and recorded. The randomization code and seeds are archived with the AEA RCT Registry dataverse upon study completion. Five sequential steps:

**Step 1 — Tournament assignment.** From ~100 recruits, randomly assign ~70 to active play and ~30 as backups. Stratified by group (tourist/resident) to ensure ~50/50 balance among active players.

**Step 2 — Team assignment.** 70 active players → 35 teams of 2. Approximately 23 mixed teams (~65%) and 12 homogeneous teams (~35%). Randomization is stratified at the individual level by gender and volleyball ability (above/below median within each group). Men and women are randomized separately: male tourists and male residents are assigned to teams independently of female tourists and female residents. This produces four team gender types (male-male, female-female, male-female, female-male) distributed across mixed and homogeneous arms. Team-level gender composition is not constrained but is recorded and included as a covariate.

**Step 3 — Incentive assignment.** Within each team type (mixed/homogeneous), randomly assign half to Team Pay and half to Individual Pay.

**Step 4 — Match schedule.** Bollobás pairing method (random simple regular graph; see Lowe, 2021, for implementation details) generates ~8 matches per team. This creates random variation in the proportion of outgroup opponents (Prop.Oth.Group.Opponents), enabling estimation of adversarial contact effects. Prop.Oth.Group.Opponents is defined by the initial randomized match schedule, not by realized opponents. If a match is forfeited or a substitute replaces the scheduled opponent, the schedule-assigned opponent composition is retained for analysis (intent-to-treat). Realized opponent composition is used in a sensitivity check.

**Step 5 — Backup priority.** Randomized priority numbers within each group (tourist/resident). High-priority backups (ranks 1–10) are called first as substitutes; low-priority backups (ranks 11–30) are called last. Team composition type (mixed or homogeneous) is never altered during the tournament. Substitutes are always drawn from the same group as the absent player, preserving the original assignment. If no same-group backup is available, the match is forfeited rather than played with an altered team composition. All participants are analyzed according to their original random assignment (intent-to-treat).

### 4.2 Balance Verification

We will report a balance table (Appendix Table A1) comparing baseline covariates across treatment arms (mixed vs. homogeneous vs. high backup vs. low backup). A joint F-test of all baseline covariates on MixedTeam will test omnibus balance. Significant imbalance ($p < 0.05$) on any single covariate will be noted; the covariate will be included as a control in all specifications regardless.

## 5. Data Collection

### 5.1 Baseline Survey (Days 7–12)

Administered in person at the tournament venue in the participant's preferred language.

**Demographics:** Age, gender, nationality, education, occupation, length of stay (tourists), years of residence (residents), prior visits to the Canary Islands (tourists).

**Social network (roster method):** Participants view a photo roster of all other recruits in their tournament and check boxes for each person: "I don't know this person," "I know this person," "This person is a friend," "This person is a close friend." **Baseline outgroup friendships** = count of outgroup members marked as "friend" or "close friend."

**Attitudes:** 5 items on attitudes toward the outgroup (tourists toward residents or vice versa), using a 7-point Likert scale. Adapted from Purper and Araña (in press).

**Language proficiency:** Self-assessed ability to communicate with a typical member of the other group, on a 1–5 scale (1 = no ability, 5 = fluent). Also: languages spoken.

**Group identifiability check:** Each participant views photos of 10 randomly selected other recruits (5 tourists, 5 residents) and classifies each as "tourist" or "resident." Accuracy rate is recorded. This validates the assumption that group membership is naturally identifiable (Allport's conditions require awareness of outgroup membership). Administered to **all** participants.

### 5.2 Volleyball Ability Assessment (Days 7–12)

Three standardized drills, each scored quantitatively:

1. **Serve accuracy:** 10 serves aimed at a marked target zone on the opposite court. Score = count landing in zone (0–10).
2. **Pass/reception quality:** 10 tosses by assessor; participant passes to a setter position. Each rated by assessor on a 1–3 scale (1 = poor, 2 = acceptable, 3 = good). Score = sum (10–30).
3. **Setting consistency:** 10 sets aimed at a target position. Score = count landing within 1 meter of target (0–10).

**Volleyball ability index** = equally-weighted mean of the 3 drill scores, each standardized (z-scored) within the tournament.

**Inter-rater reliability protocol for pass/reception drill:** Two independent raters score a random 20% subsample of pass drills. Intra-class correlation coefficient (ICC) is computed. If ICC $< 0.70$, raters are retrained and the subsample is re-scored before proceeding. The ICC is reported in the paper appendix.

### 5.3 Endline-1: Trading Task (Days 35–44, In Person)

Participants receive 2 mismatched beach items during the final week of tournament play (day 35). They have 10 days (days 35–44, overlapping with the last 6 tournament days and 4 days post-tournament) to find trading partners among all tournament participants. A randomized cross-group bonus (€2 or €5) is paid for each successful trade with an outgroup member. Trades are recorded by scanning numbered item cards.

**Outcome:** Binary indicator for whether the participant completed at least one cross-group trade. For each trade, we record **whether the trading partner was the participant's teammate or a non-teammate** outgroup member.

**Sensitivity analysis:** Code the trading outcome as 1 only if the participant traded with a *non-teammate* outgroup member (excluding teammate trades). This addresses the concern that mixed-team participants may have exchanged contact information with their outgroup teammate during matches, giving them a built-in trading partner that conflates willingness-to-trade with convenience/opportunity. Report the fraction of cross-group trades with teammates vs. non-teammates, by treatment arm.

**Construct clarification.** The cross-group bonus incentivizes outgroup trading, so the outcome measures willingness to initiate cross-group economic exchange at a given bonus level — not unconditional willingness. The bonus amount (€2 vs. €5) is included as a covariate in the trading analysis. The bonus × MixedTeam interaction is pre-registered as exploratory.

**Sequencing note.** The Endline-1 trading window (Days 35–44) precedes Endline-2 (Days 48–60). Participants who successfully trade with outgroup members during Endline-1 experience additional positive cross-group interaction beyond the original treatment. Since mixed-team participants may trade cross-group at higher rates, this creates a potential post-treatment mediator: Endline-2 outcomes could partly reflect the trading experience rather than solely the team assignment. We acknowledge this as a limitation. As a sensitivity analysis, we estimate Equation (1) on Endline-2 outcomes controlling for whether the participant completed a cross-group trade. This is a "bad control" in the Angrist-Pischke sense (conditioning on a post-treatment variable biases the treatment effect), but the comparison of controlled vs. uncontrolled estimates is informative about the mediating role of the trading task. The Endline-2-only index (Section 12.3) also addresses this by excluding the trading component.

### 5.4 Endline-2: Behavioral Exercises (Days 48–60, Online for All)

Administered via Qualtrics with unique participant links, for **all participants** (tourists and residents alike), to ensure uniform measurement conditions and eliminate measurement-mode confounds.

**Social interaction exercise.** Participant scrolls through a photo roster of all tournament participants. For each person, they indicate: (a) "I would want to spend time with this person" and (b) "I consider this person a friend." Outcomes: count of outgroup members selected in each category.

**Team formation exercise.** Participant chooses 1 teammate for a hypothetical future match. Two versions: (a) with a €50 prize for winning; (b) without a prize. The choice set includes names/photos of tournament participants from both groups. Outcome: binary indicator for whether an outgroup member is chosen (one per version).

**Voting exercise.** Participant ranks players from 4 randomly selected other teams for a desirable prize (€30 beach excursion or restaurant voucher). The 4 teams are randomly drawn subject to the constraint that **at least 1 team contains an ingroup member and at least 1 team contains an outgroup member**, ensuring that ingroup favoritism is measurable. Borda count scoring.

Outcome: degree of ingroup favoritism in vote rankings (difference in average rank assigned to ingroup vs. outgroup candidates).

**Trust game.** Standard trust game with €10 endowment. Participant sends €0–10 to a partner; amount is tripled; partner returns €0 to 3× the received amount. Each participant plays with 2 partners: 1 tourist and 1 resident, both **randomly drawn** from the pool of participants in other tournaments/locations who have completed Endline-2 (to measure generalized outgroup trust, not trust toward specific tournament acquaintances). Partner matching is random and anonymous — participants see only the partner's group label ("a tourist from a different tournament" or "a local resident from a different tournament"). Both sender and returner decisions are real and binding (double-blind protocol). Implemented as a strategy-method game online: each participant makes both sender and returner decisions; one role is randomly selected for payment. Real payoffs via bank transfer within 2 weeks. **Primary outcome:** amount sent to the outgroup partner (sender role). **Exploratory outcome:** amount returned to the outgroup partner (returner role). Partners from different tournaments may differ systematically (different locations, waves, treatment arms); this noise attenuates treatment effect estimates toward zero.

**Stated trust question.** WVS-style question: "Generally speaking, would you say that most [tourists/residents] can be trusted, or that you need to be very careful in dealing with them?" Rated on a 1–10 Likert scale.

**Post-tournament group identifiability verification.** All participants answer: "During the tournament, did you know whether your teammate was a tourist or a local resident?" (Yes/No/Unsure). This validates the operational assumption that group membership was salient during the intervention. Reported in the paper.

### 5.5 Match-Level Data

For each match: date, teams, scores, individual point tallies (for Individual Pay calculation), attendance (which players showed up), substitute assignments. Collected by tournament logistics RAs (separate from endline RAs for blinding purposes).

---

## 6. Outcome Variables

### 6.1 Primary Outcome: Omnibus Cross-Group Behavior Index

The primary inference target, following Lowe (2021, Section IIID) and Kling, Liebman, and Katz (2007).

**Construction:**

1. For each of the 7 components below, compute the **control group mean** ($\mu_c$) and **standard deviation** ($\sigma_c$), where the control group comprises zero-match backups — low-priority backups who were never called as substitutes during the tournament. Zero-match backups provide a clean reference distribution uncontaminated by partial treatment exposure.
2. Winsorize continuous components at the 1st and 99th percentiles (Section 10.2) before standardization.

3. Standardize each component: $z_k = (y_k - \mu_c) / \sigma_c$
4. Omnibus index = $(1/7) \times \Sigma_{k=1}^{7} z_k$

**Control group note.** Zero-match backups provide the cleanest reference distribution. As a sensitivity analysis, we recompute standardization using all low-priority backups by assignment status (ITT within the backup group), including those who substituted into one or two matches, and using the full-sample standard deviation. We report the distribution of actual matches played by low-priority backups and the fraction who substituted onto mixed vs. homogeneous teams.

**Components:**

| # | Component | Source | Type | Range |
|---|---|---|---|---|
| 1 | Want-to-interact count (outgroup) | Endline-2 | Count | 0 to N_outgroup |
| 2 | Friends count (outgroup) | Endline-2 | Count | 0 to N_outgroup |
| 3 | Outgroup chosen for team (stakes) | Endline-2 | Binary | {0, 1} |
| 4 | Outgroup chosen for team (no stakes) | Endline-2 | Binary | {0, 1} |
| 5 | Cross-group trade completed | Endline-1 | Binary | {0, 1} |
| 6 | Trust game amount sent to outgroup | Endline-2 | Continuous | €0–10 |
| 7 | Stated generalized trust (outgroup) | Endline-2 | Likert | 1–10 |

All components are oriented so that higher values indicate more cross-group prosocial behavior.

### 6.2 Secondary Outcomes

**The 7 omnibus index components** (Section 6.1), analyzed separately using the same primary specification (Equation 1).

**Voting/favoritism** (8th secondary outcome, not in the omnibus index). Degree of ingroup favoritism in the voting exercise (Section 5.4), measured as the difference in average Borda rank assigned to ingroup vs. outgroup candidates (reverse-coded so higher = less favoritism = more prosocial). This outcome is excluded from the omnibus index because its continuous-difference metric (Borda rank differential) has different distributional properties from the other 7 components (which are counts, binary indicators, or bounded Likert/currency amounts). Including it would require additional distributional assumptions for valid index aggregation. It is analyzed as a standalone secondary outcome using Equation (1) and is subject to the same Benjamini-Hochberg FDR correction as the other secondary outcomes (Family 2, now 8 outcomes).

### 6.3 Outcome Families for Multiple Testing

| Family | Outcomes | Correction | Rationale |
|---|---|---|---|
| 1 (Primary) | Omnibus index | None (single test) | Pre-registered primary outcome |
| 2 (Secondary) | 8 outcomes: 7 index components + voting/favoritism | Benjamini-Hochberg FDR, q = 0.10 | Multiple related outcomes |
| 3 (Confirmatory heterogeneity) | MixedTeam × Group, MixedTeam × Gender, MixedTeam × Language proficiency | Bonferroni, $\alpha/K = 0.05/3 \approx 0.017$ per test, two-sided | Theoretically motivated moderators |
| 4 (Exploratory heterogeneity) | MixedTeam × {Age, Ability, Stay, Baseline friends, Win/loss} | BH-FDR, q = 0.10 | Data-driven exploration |
| 5 (Exploratory specifications) | Equations 3 and 4 | No formal correction; flagged as exploratory | Underpowered / different estimand |

---

## 7. Estimation Strategy

### 7.1 Primary Specification — Equation (1): Effects of Contact Type

Estimated on **active tournament players only** (N ≈ 560 at 0% attrition).

$$y_{igl} = \alpha_{gl} + \beta \cdot MixedTeam_{igl} + \gamma \cdot Prop.Oth.Group.Opponents_{igl} + \eta \cdot X_{igl} + \varepsilon_{igl}$$

| Term | Definition |
|---|---|
| $y_{igl}$ | Outcome for participant i from group g in tournament l (standardized by control group μ and σ) |
| $\alpha_{gl}$ | Group × tournament fixed effects (absorb all location-wave-group-level variation) |
| $\beta$ | **Primary parameter of interest:** effect of having an outgroup teammate (collaborative contact) |
| $\gamma$ | Effect of proportion of outgroup opponents across all matches (adversarial contact) |
| $X_{igl}$ | Baseline covariates (see below) |
| $\varepsilon_{igl}$ | Error term |

**Baseline covariates (X_igl):** 1. Age (years) 2. Gender (female = 1) 3. Volleyball ability index (standardized within tournament) 4. Baseline outgroup friendships (count from roster method)

5. Missing social network survey indicator (= 1 if participant did not complete roster) 6. Stratification variables used in randomization (gender composition stratum, ability stratum) 7. Language proficiency (1–5 self-assessment)

**Standard errors:** - **Primary:** Clustered at the team level (280+ clusters across 8 tournaments) - **Robustness 1:** Wild cluster bootstrap at the tournament level (8 clusters), following Cameron, Gelbach, and Miller (2008) - **Robustness 2:** Randomization inference (Fisher's exact test, 5,000 permutations of treatment assignment within strata)

**Interpretation of β.** Because MixedTeam $\in \{0, 1\}$, $\beta$ is the average treatment effect of having 1 outgroup teammate (vs. 0) on the outcome, measured in control-group standard deviation units. Although H1 predicts $\beta > 0$, we use a **two-sided test** at $\alpha = 0.05$ to allow detection of adverse effects ($\beta < 0$), consistent with best practice for pre-registered experiments.

**Identification of γ.** MixedTeam and Prop.Oth.Group.Opponents are both included in Equation (1). They are potentially correlated because the match schedule (Step 4) is generated conditional on team assignments. However, $\gamma$ is identified from **within-team-type variation** in the match schedule: the Bollobás pairing creates random variation in opponent composition conditional on MixedTeam. We will report the correlation between MixedTeam and Prop.Oth.Group.Opponents from the realized randomization. As a robustness check, we estimate Equation (1) dropping Prop.Oth.Group.Opponents to verify that $\beta$ is stable (Section 12.1).

### 7.2 Secondary Specification — Equation (2): Program Evaluation

Estimated on the **full sample including backups** (N ≈ 800 at 0% attrition).

$$y_{igl} = \alpha_{gl} + \varphi_1 \cdot \text{Homog.Team}_{igl} + \varphi_2 \cdot \text{Mixed.Team}_{igl} + \varphi_3 \cdot \text{High.Backup}_{igl} + \eta \cdot X_{igl} + \varepsilon_{igl}$$

| Term | Definition |
|---|---|
| Omitted category | Low-priority backups (~1–2 matches, approximate control) |
| $\varphi_1$ | Effect of homogeneous team (vs. control) |
| $\varphi_2$ | Effect of mixed team (vs. control) |
| $\varphi_3$ | Effect of high-priority backup (~3–4 matches, short-term contact) |
| $\varphi_2 - \varphi_1$ | Additional effect of mixed vs. homogeneous teams |

Standard errors clustered at team level. Backups assigned to a "solo team" cluster for clustering purposes.

### 7.3 Exploratory Specification — Equation (3): Incentive Interaction

**Pre-registered as EXPLORATORY.** The MDE for the interaction term $\beta_2$ is approximately 0.40–0.44σ for individual outcomes and ~0.25–0.28σ for the omnibus index — far larger than plausible effect sizes. A significant $\beta_2$ would be informative; a null result is uninformative.

$$y_{igl} = \alpha_{gl} + \beta_1 \cdot \text{MixedTeam}_{igl} + \beta_2 \cdot \text{MixedTeam}_{igl} \times \text{TeamPay}_{igl} + \beta_3 \cdot \text{TeamPay}_{igl} + \eta \cdot X_{igl} + \varepsilon_{igl}$$

$\beta_2 > 0$ would indicate that strengthening financial reward interdependence (team pay) enhances the collaborative contact effect on cross-group prosocial behavior.

### 7.4 Exploratory Specification — Equation (4): Dyadic Generalization

**Pre-registered as EXPLORATORY.** Estimated at the participant-pair level, following Lowe (2021, equations 3–4).

$$y_{ij} = (\alpha_{jgl} \times \text{MixedTeam}_{igl}) + \beta_1 \cdot \text{Teammate}_{ij} + \beta_2 \cdot \text{Friend.of.Teammate}_{ij} + \zeta \cdot X_{ij} + \varepsilon_{ij}$$

| Term | Definition |
|---|---|
| $y_{ij}$ | Outcome for participant i toward participant j |
| Teammate_ij | Indicator: j was i's teammate |
| Friend.of.Teammate_ij | Indicator: j was a baseline friend of i's teammate |
| $\beta_1$ | Direct teammate effect |
| $\beta_2$ | Network spillover effect |

Standard errors two-way clustered at the (i, j) level, following Cameron and Miller (2015) for dyadic regressions, to account for correlation among all pairs sharing the same i or the same j. The j-specific fixed effects absorb j-level means but not the within-j correlation structure. If two-way clustering is computationally infeasible, we fall back to i-level clustering (which is conservative given the j fixed effects) and note this in the paper. The j-specific fixed effects interacted with MixedTeam_igl imply approximately 2 × N_outgroup_per_tournament fixed effects per tournament-group cell (e.g., ~70 for a 35-outgroup-member cell), which is computationally feasible with standard panel estimators but consumes degrees of freedom. We verify that the effective sample size per cell exceeds 20 observations; if any cell falls below this threshold, we collapse cells or drop the cell and note this deviation.

### 7.5 Heterogeneity Analyses

All heterogeneity tests interact MixedTeam with the moderator variable in Equation (1). The moderator and the MixedTeam × Moderator interaction are added to the baseline specification.

**Confirmatory (Bonferroni-corrected, K = 3, reject at $\alpha/K = 0.05/3 \approx 0.017$ per test, two-sided):**

1. **Group membership** (tourist = 1, resident = 0): Tests whether collaborative contact effects are symmetric for hosts and guests.
2. **Gender** (female = 1): Tests whether effects differ by gender, relevant because Lowe (2021) studied men only.
3. **Language proficiency** (above/below median baseline ability to communicate with the other group): Tests whether treatment effects are driven by communication ease rather than prejudice reduction. If effects are concentrated among high-proficiency pairs, language barriers

may confound the contact mechanism. If effects are similar across proficiency levels, communication friction is unlikely to explain the results.

**Exploratory (Benjamini-Hochberg FDR, q = 0.10):**

4. **Age** (median split: above/below median age in the tournament)
5. **Volleyball ability** (above/below median ability index)
6. **Length of stay** — tourists only (above/below median stay length)
7. **Baseline outgroup friendships** (any vs. none at baseline)
8. **Match outcomes** (above/below median win proportion)

---

## 8. Attrition and Missing Data

### 8.1 Attrition Testing

At each endline, we estimate:

$$Attrit_{igl} = \delta_0 + \delta_1 \cdot MixedTeam_{igl} + \delta_2 \cdot Prop.Oth.Group.Opponents_{igl} + \delta_3 \cdot X_{igl} + \nu_{igl}$$

where Attrit = 1 if participant i did not complete the endline. Insignificant $\delta_1$ supports the assumption that attrition is not selective on treatment (as in Lowe, 2021, Appendix Table A5).

**Group-stratified attrition tests.** Because tourists and residents have different expected attrition rates (~20% vs. ~10% at Endline-1; ~35% vs. ~15% at Endline-2), the attrition regression is estimated **separately for tourists and residents** in addition to the pooled specification. This allows detection of differential attrition that may be masked in the pooled sample.

### 8.2 Bounding Strategies

All pre-specified and applied to the primary outcome (omnibus index) and all secondary outcomes:

1. **Lee (2009) bounds.** Trim the treatment arm with lower attrition to equalize completion rates. Report upper and lower bounds on β. **Computed separately for tourists and residents** (group-stratified Lee bounds) in addition to pooled bounds.
2. **Horowitz-Manski (2000) worst-case bounds.** Assign best possible outcomes to attriters in one arm and worst to the other. Expected to be uninformative (too wide) but included for completeness.
3. **Inverse probability weighting (IPW).** Model P(completion | baseline covariates) using a logit model that includes: age, gender, group, volleyball ability, stay length, baseline outgroup friendships, and **group × MixedTeam interactions** in the attrition model. Weight complete-case observations by $1/\hat{P}$(completion). Report IPW-adjusted β alongside unweighted.
4. **Baseline comparison.** Compare baseline characteristics of completers vs. attriters, separately by treatment arm and by group. Report as Appendix Table A-Attrition.

### 8.3 Missing Data

- **Missing baseline covariates:** Impute at the group × tournament cell mean. Include a missing-data indicator for each imputed covariate.
- **Missing outcomes:** Listwise deletion for the primary analysis. IPW as robustness (Section 8.2, item 3).
- **Partially missing omnibus index:** If a participant completes some but not all 7 component exercises, the omnibus index is computed over the available components (adjusting K accordingly). Sensitivity: restrict to participants with all 7 components.

### 8.4 Contingency Plan for Low Endline-2 Response

**Pre-registered decision rule:** If <60% of Wave 1 participants complete Endline-2 within 14 days of link distribution, the following adaptive measures are triggered for all subsequent waves:

1. Phone/WhatsApp follow-up by RAs at days 7 and 14 after link distribution
2. Increase Endline-2 completion bonus from €5 to €10
3. Extend response window from 14 to 21 days

This contingency is pre-registered to avoid post-hoc changes to the data collection protocol. The decision rule is based on Wave 1 response rates only; it does not involve examining treatment effects.

**Statistical implications.** If the contingency is triggered, Waves 2+ will have a different protocol (higher bonus, longer window, phone follow-up). All waves are **pooled in the primary analysis regardless** of whether the contingency was triggered. A "post-contingency wave" indicator is included as a covariate in a robustness check to test for protocol-driven differences. We note that the increased completion bonus (€10 vs. €5) could affect trust game behavior if participants feel wealthier; this is acknowledged as a limitation but considered second-order given the small bonus magnitude relative to the trust game endowment (€10).

---

## 9. Power Analysis

### 9.1 Parameters

| Parameter | Value | Source |
|---|---|---|
| Treatment | Binary: MixedTeam ∈ {0, 1} | 2v2 teams |
| Allocation | 65% mixed, 35% homogeneous | Following Lowe (2021) |
| Var(MixedTeam) | 0.2275 | Bernoulli variance |
| ICC (team) | ρ = 0.05 | Conservative; m = 2 players per team |
| DEFF | 1.05 | 1 + (2−1)(0.05) |
| Covariate $R^2$ | 0.15 | Baseline covariates |
| σ_ε | 0.922 | √(1 − 0.15) |
| α | 0.05 (two-sided) | Standard |
| Power | 0.80 | Standard |

### 9.2 MDE Formula

```
MDE = 2.80 × 0.922 / √(N_analyzed × 0.2275 / 1.05) = 5.54 / √N_analyzed
```

### 9.3 MDE Tables

**Individual outcomes:**

| Recruited N | Attrition | Analyzed N | MDE (σ) |
|---|---|---|---|
| 800 | 0% | 800 | 0.196 |
| 800 | 15% | 680 | 0.212 |
| 800 | 20% | 640 | 0.219 |
| 800 | 25% | 600 | 0.226 |
| **800** | **35%** | **520** | **0.243** |
| 1,200 | 15% | 1,020 | 0.173 |
| 1,200 | 20% | 960 | 0.179 |
| 1,200 | 35% | 780 | 0.198 |

**Omnibus index** (variance reduction factor = 0.40, assuming K = 7 components with ρ_out = 0.3):

| Recruited N | Attrition | Analyzed N | MDE individual | MDE omnibus |
|---|---|---|---|---|
| 800 | 0% | 800 | 0.196 | 0.124 |
| 800 | 15% | 680 | 0.212 | 0.134 |
| 800 | 20% | 640 | 0.219 | 0.138 |
| **800** | **35%** | **520** | **0.243** | **0.154** |
| 1,200 | 15% | 1,020 | 0.173 | 0.110 |
| 1,200 | 20% | 960 | 0.179 | 0.113 |
| 1,200 | 35% | 780 | 0.198 | 0.125 |

**Active players only — Equation 1 (primary confirmatory test, N ≈ 560):**

| Active players | Attrition | Analyzed N | MDE individual | MDE omnibus |
|---|---|---|---|---|
| 560 | 0% | 560 | 0.234 | 0.148 |
| 560 | 15% | 476 | 0.254 | 0.161 |

*Note.* The full-sample MDEs above (N ≈ 800) apply to Equation 2 (program evaluation). Equation 1 (primary confirmatory test) is estimated on active tournament players only (N ≈ 560 at 0% attrition). MDEs assume ICC = 0.05, $R^2$ = 0.15, 65/35 allocation, and team-level clustering (design effect = 1.05). Omnibus index MDEs apply the variance reduction factor of 0.40.

### 9.4 Interpretation

- **At 0–20% attrition (likely for Endline-1):** The omnibus MDE is 0.12–0.14σ, comparable to Lowe's (2021) ex-post MDE of 0.14σ. Sufficient to detect Lowe-scale effects (0.23σ) with >95% power and effects ≥0.15σ with ~80% power.
- **At 35% attrition (plausible for Endline-2):** The omnibus MDE rises to 0.154σ. The design can still detect effects ≥0.16σ but loses power for effects in the 0.10–0.15σ range. The most likely scenario (30–35% Endline-2 attrition) yields an omnibus MDE of ~0.14–0.15σ.
- **Comparison:** Lowe (2025) meta-analysis reports an average clean contact effect of 0.08σ across 41 experiments. Our design cannot detect effects this small. However, Lowe's own collaborative contact effect (0.23σ) is well within our detectable range.

### 9.5 MDE Sensitivity to Intra-Cluster Correlation

With m = 2 players per team, the design effect DEFF = 1 + (m−1)ρ is relatively insensitive to ICC. Teammates who play together for 3+ weeks may have more correlated outcomes than assumed at ρ = 0.05. Sensitivity:

| ICC (ρ) | DEFF | MDE individual (N=800, 0% attrition) | MDE omnibus |
|---|---|---|---|
| 0.05 | 1.05 | 0.196σ | 0.124σ |
| 0.10 | 1.10 | 0.201σ | 0.127σ |
| 0.15 | 1.15 | 0.205σ | 0.130σ |
| 0.20 | 1.20 | 0.210σ | 0.133σ |

Even at ρ = 0.20 (an implausibly high ICC for heterogeneous 2-person teams), the omnibus MDE increases by only 0.009σ — from 0.124σ to 0.133σ. The qualitative power conclusions are robust to ICC assumptions. The primary ρ = 0.05 assumption is conservative relative to within-pair correlations typically observed in comparable cluster-randomized experiments with small clusters.

### 9.6 MDE for the Interaction Term (Equation 3)

The MixedTeam × TeamPay interaction requires ~4× the sample for equivalent power:

| Outcome | Attrition | MDE for interaction |
|---|---|---|
| Individual | 0% | ~0.39σ |
| Individual | 20% | ~0.44σ |
| Omnibus | 0% | ~0.25σ |
| Omnibus | 20% | ~0.28σ |

These MDEs exceed plausible interaction effect sizes, confirming that Equation (3) is exploratory.

---

## 10. Variable Construction Rules

### 10.1 Outcome Orientation

All outcomes are coded so that **higher values = more cross-group prosocial behavior**. The voting/favoritism outcome is reverse-coded (less ingroup favoritism = higher value).

### 10.2 Outlier Treatment

- **Continuous outcomes** (trust game amount, want-to-interact count, friends count): winsorize at the 1st and 99th percentiles of the full sample distribution.
- **Sensitivity:** Winsorize at the 5th and 95th percentiles.
- **Binary outcomes** (team formation, trading): No outlier treatment needed.

### 10.3 Trading Task: Bonus Amount

The randomized bonus (€2 or €5) is included as a covariate in all specifications involving the trading outcome. The number of **days available for trading** (days between item distribution and departure or end of the trading window, whichever comes first) is also included as a covariate in all trading-task specifications. The outcome measures willingness to trade cross-group at the given bonus level. The bonus × MixedTeam interaction is exploratory (not formally corrected).

### 10.4 Exclusion Rule

**Main analysis:** Exclude active tournament players who attended <3 of their ~8 scheduled matches. This threshold was chosen **before seeing any data** and is based on a dosage argument: participants with <3 matches (37.5% attendance) received insufficient collaborative contact for the treatment mechanism to operate meaningfully — fewer than 3 matches provides less than ~6 hours of cooperative interaction, below the minimum dose observed in effective contact interventions (Lowe, 2021, used ~15 hours across the season).

**Sensitivity analyses:** - Threshold at <2 matches (more inclusive) - Threshold at <4 matches (more restrictive) - Include all active players regardless of attendance (ITT)

Backups are not subject to the attendance exclusion; they are included in Equation (2) based on their actual number of matches played.

### 10.5 Index Construction

**Primary index:** Equally-weighted mean of 7 standardized components (Section 6.1).

**Robustness index:** Inverse-covariance-weighted index following Anderson (2008), using the **control group (low-priority backups) covariance matrix** for weighting. This weights components inversely by their covariance, giving more weight to components that contain independent information.

---

## 11. Balance and Validity Checks

### 11.1 Balance Table (Appendix Table A1)

Baseline covariates compared across 4 arms (mixed team, homogeneous team, high-priority backup, low-priority backup):

- Age
- Gender (% female)
- Nationality distribution (% German, % British, % Scandinavian, % Spanish, % other)
- Volleyball ability index
- Baseline outgroup friendships (count)
- Language proficiency (1–5)
- Attitudes toward outgroup (mean of 5 Likert items)
- Photo classification accuracy (% correct in group identifiability check)
- Length of stay in days (tourists only; reported separately and excluded from the pooled joint F-test since it is undefined for residents)

Joint F-test of all covariates on MixedTeam reported (excluding tourist-only variables). Individual t-tests for each covariate.

### 11.2 Group Identifiability Check

**Baseline photo task:** Report mean classification accuracy across all participants. If accuracy is significantly below 70%, discuss implications for the contact mechanism (participants may not have been aware of group membership during matches).

**Post-tournament verification:** Report the percentage of participants who answered "Yes" to "Did you know whether your teammate was a tourist or a local resident?" If <80% report "Yes," this weakens the treatment interpretation and is discussed as a limitation.

### 11.3 Randomization Verification

Verify that the implemented randomization matches the computer-generated assignment. Report any deviations (e.g., last-minute team changes due to no-shows) and their frequency.

---

## 12. Robustness Checks

### 12.1 Specification Robustness

| Check | Description |
|---|---|
| No controls | Equation (1) without X_igl (only α_gl and treatment variables) |
| Drop adversarial contact | Equation (1) without Prop.Oth.Group.Opponents (verifies β stability) |
| Extended controls | Add baseline attitudes, language proficiency, nationality dummies |
| Without fixed effects | Replace α_gl with separate group and tournament dummies |
| Control for Endline-1 trading | Equation (1) on Endline-2 outcomes adding cross-group trade indicator as control (see Section 5.3 sequencing note; acknowledges bad-control caveat) |

### 12.2 Inference Robustness

| Check | Description |
|---|---|
| Wild cluster bootstrap | Tournament-level clustering (8 clusters), Cameron et al. (2008) |
| Randomization inference | Fisher's exact test, 5,000 permutations within strata |
| Conley spatial SEs | If within-tournament spatial correlation is suspected |

### 12.3 Index Robustness

| Check | Description |
|---|---|
| Anderson (2008) ICW index | Inverse-covariance-weighted instead of equal-weight |
| Drop each component | 7 leave-one-out indices to check that no single component drives results |
| Endline-1-only index | Index using only Component 5 (trading) — no Endline-2 attrition |
| Endline-2-only index | Index using Components 1–4, 6–7 — for the online subsample |

### 12.4 Sample Robustness

| Check | Description |
|---|---|
| Attendance thresholds | Exclusion at <2, <3 (primary), <4 matches |
| Full ITT | Include all active players regardless of attendance |
| Exclude re-participating residents | Contingency for imperfect enforcement of the screening protocol — applies only if any residents are detected ex post to have participated in more than one wave despite baseline screening |
| Tourists only | Estimate Equation (1) on the tourist subsample |
| Residents only | Estimate Equation (1) on the resident subsample |
| Post-contingency indicator | Add wave-level indicator for whether contingency protocol (Section 8.4) was active |
| Exclude relaxed-criterion tourists | If the recruitment fallback (≥18-day threshold) was triggered at any location-wave, re-estimate Equation (1) excluding tourists recruited under the relaxed criterion to verify that results are robust to the stricter ≥21-day inclusion |
| Exclude short-trading-window participants | Exclude participants with fewer than 3 trading days from the trading-task outcome; re-estimate Equation (1) on the trading component and the omnibus index |
| Alternative control-group standardization | Standardize using all low-priority backups by assignment status (ITT), including those who substituted into 1–2 matches |

### 12.5 Consent Languages

Informed consent is available in Spanish, English, German, Swedish, and Norwegian. Finnish-speaking participants (if any) are offered the English version. This coverage reflects the dominant

tourist nationalities in the Canary Islands. Consent language availability is documented in the ethics protocol.

---

## 13. Data Management

**Storage.** Raw data are stored on ULPGC's institutional secure server with automated daily backups. Access restricted to PIs and authorized RAs.

**De-identification.** Participant identifiers are replaced with study IDs immediately after data collection for each endline. The linking key (study ID ↔ personal data) is stored separately in an encrypted file accessible only to PIs.

**Archiving.** Upon publication, de-identified analysis datasets and all Stata/R analysis code are posted to the AEA RCT Registry dataverse. Randomization code and seeds are included. Personal data (names, contact information, photos used in roster tasks) are **not** archived.

**GDPR/LOPDGDD compliance.** Detailed in the ethics protocol. Participants can withdraw consent and request data deletion at any point before de-identification. After de-identification, withdrawal removes the linking key but de-identified data remain in the analysis dataset (as pre-registered).

---

## Appendix A: Planned Tables and Figures

### Tables

| # | Content | Specification |
|---|---|---|
| 1 | Outcome exercise descriptions | Descriptive |
| 2 | Social interaction: want-to-interact and friends | Eq. (1) |
| 3 | Team formation: with and without stakes | Eq. (1) |
| 4 | Voting: ingroup favoritism | Eq. (1) |
| 5 | Trading: cross-group trade | Eq. (1) + bonus covariate |
| 6 | Trust game: amount sent, stated trust | Eq. (1) |
| 7 | Omnibus cross-group behavior index | Eq. (1) |
| 8 | Program evaluation: mixed vs. homogeneous vs. backups | Eq. (2) |
| A1 | Balance checks | t-tests + joint F-test |
| A2 | Summary statistics | Full sample + by group |
| A3 | Attrition analysis | Eq. in Section 8.1, pooled + group-stratified |
| A4 | Lee (2009) bounds | Pooled + group-stratified |
| A5 | Heterogeneity: group × treatment, gender × treatment | Eq. (1) + interactions |

### Figures

| # | Content |
|---|---|
| 1 | Experiment design and timeline |
| 2 | Collaborative and adversarial effects on social interaction (2-bin + binscatter) |
| 3 | Generalization: effects on non-teammates |
| 4 | Omnibus index: collaborative and adversarial (2-bin + binscatter) |
| 5 | Program evaluation bar chart (control vs. homogeneous vs. mixed) |

## Appendix B: McKenzie (2012) Checklist Verification

| # | Element | PAP Section |
|---|---|---|
| 1 | Sample description | Section 3 |
| 2 | Key data sources | Section 5 |
| 3 | Hypotheses | Section 1.2 |
| 4 | Variable construction | Sections 5, 6, 10 |
| 5 | Treatment effect equation | Section 7 |
| 6 | Multiple outcomes strategy | Section 6.3 |
| 7 | Survey attrition procedures | Section 8 |
| 8 | Limited variation solutions | Section 10.4 (exclusion rules) |
| 9 | Theoretical model | Section 2.3 (Allport's conditions) |
| 10 | Registry filing | AEA RCT Registry |

*Pre-Analysis Plan v1.7 — 2026-02-22 — To be filed with AEA RCT Registry before data collection begins.*