

Online Appendix: Endogenous Surplus Retention in Peer-Based Team Incentives

This Online Appendix collects technical derivations and estimation details referenced in the main text as Online Appendix A–E. It is supplementary material and is not part of the main manuscript.

Appendix A. Feasibility note on the Shapley value

Shapley-type allocation requires 2^N coalition evaluations (prohibitive for $N > 25$) and presumes observability of sub-coalition surplus that team production lacks; the fixed denominator evaluates in $O(N^2P)$ from observed votes. The comparison is a feasibility note, not a normative one: the two rules answer different questions, and Shapley’s information requirements are not met in the setting considered here. We therefore do not pursue it as a benchmark.

Appendix B. Estimation of the recognition–effort elasticity (Prop. 11)

Within each class we estimate the recognition–effort elasticity γ from

$$\ln w_i = \ln k + \gamma \ln a_i + \varepsilon$$

(class-demeaned OLS, a_i the instructor anchor; 1,216 submissions with $w_i, a_i > 0$). Class-level mean $\bar{\gamma} = 0.41$ (median 0.39; range 0.05–0.78; bootstrap 95% CI [0.32, 0.50]); pooled with class fixed effects 0.37 (SE 0.016; 95% CI [0.34, 0.40]; rank regression 0.42), positive in all 20/20 classes — above the Prop. 11 threshold by a factor of about 46 (0.009 at the CHDYPY N and G). The estimate is conservative: it inherits the noise of $\rho_{\text{pool}} = 0.439$ rather than correcting for it. Note that γ here is the recognition–effort *elasticity* — the exponent of Prop. 11’s signal model $z_i \propto a_i^\gamma$ — estimated directly by regression; it is a different quantity from the rank correlation ρ reported in §5.2 (monotone association), and we do not recover one from the other.

Appendix C. The single raw-cap violation under the coverage-adjusted bound

The raw structural bound $Z_i < 1/N$ holds in 19/20 CHDYPY cohorts and peer-grading; it fails only in Exp.2_controlGroup_3, where one reviewee was scored 5 times instead of 3 ($Z = 0.0244 > 1/N = 0.0167$).

The coverage-adjusted bound contains it exactly:

$$\frac{w_i}{N r_{\text{design}} M} \leq \frac{r_i}{r_{\text{design}}} \cdot \frac{1}{N} \implies \frac{5}{3} \cdot \frac{1}{60} = 0.0278 > 0.0244.$$

Across all cohorts the adjusted bound is violated 0 times. The raw violation is not a counterexample but a confirmation of the general bound, made visible by broken uniformity: uneven coverage rescales, not breaks, the structural cap, which is why the scoring-slot parameter r must be an enforced design parameter rather than an incidental feature of the data.

Appendix D. Governance invariance across datasets

CHDYPY. With $\beta_j = 0.5 + K_j / \sum K$, the weighted identity holds to machine precision: $\max |s_V^F - (1 - G_V)s_V^P| = 3.5 \times 10^{-18}$. The β 's are near-constant (std 0.0024), so this dataset verifies only structural invariance, not whether weighting reorders shares. **peer-grading.** Assessor and submission IDs are disjoint (0% mapping), so K_j is unobservable, β degenerates to a constant, and the test is vacuous; we state this rather than report a number. **t4e.** No usable grader→gradee KPI chain; not tested. The stress test with full weight dispersion is delegated to the §6.6 governance simulation of the main text; neither dataset supports a claim about the behavioral version of the governance claim.

Appendix E. Simulation numerics for the conditional output comparison

Solving the fixed-denominator and equal-division first-order conditions for a symmetric team ($A = 1$, $\beta = 0.5$, $c = 1$): for $N = 10$, $G = 13.5\%$ the threshold is $\gamma^* = 0.009$; at $\gamma = 0.2$ the fixed-denominator output exceeds equal division by 58% ($N = 10$) and 95% ($N = 20$); at the empirically estimated $\gamma \approx 0.4$ (§5.2) the gain is 92% and 141%, respectively. Larger funds demand stronger signals ($G = 50\%$: threshold 0.056), the §5.7 trade-off. Closed-form thresholds match numerics to machine precision.

Appendix F. Benchmark comparison and retained-fund ownership

F.1 Peer–teacher agreement benchmark. Our signal estimates ($\rho \approx 0.44\text{--}0.60$) sit below published peer–teacher agreement ($r = 0.69$ meta-analytic across 48 studies, Falchikov & Goldfinch 2000; per-grade ≈ 0.85). The published benchmark pairs an individual’s grade with the teacher’s for the same submission, whereas our ρ correlates *received votes* (the object entering the allocation) with an independent anchor – the aggregation layer the fixed denominator is designed to handle, not a shortfall of the rule.

F.2 Retained-fund ownership (Prop. 5). The fund carries an ownership vector $\alpha_i \geq 0, \sum_i \alpha_i \leq 1$: member i ’s claim on E_{public} . With no member claim ($\alpha_i = 0$ for all i), the mechanism is immune to self-interested manipulation; with a proportional claim, a depression channel opens and calls for robust aggregation (§4.8 of the main text). The institution deciding the fund’s use is upstream of the mechanism; what the mechanism contributes is that the fund’s *size* is not a free parameter.