Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Selection, Credential Quality, and Compensatory Returns: A Course-Level Event-Study of Academic Competition Awards and Student Achievement in Chinese Higher Education

Bin Wei
Guangdong Ocean University
Jing Liu
fishinggor@gdou.edu.cn

Guangdong Ocean University

Research Article

Keywords: academic competitions, student engagement, academic achievement, compensatory returns, credential quality, quasi-experimental design

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11382787/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Selection, Credential Quality, and Compensatory Returns: A Course-Level Event-Study of Academic Competition Awards and Student Achievement in Chinese Higher Education

Abstract

Academic competition awards are a heavily institutionalised feature of Chinese undergraduate education, and they carry formal weight in admission and scholarship decisions, which makes them a high-stakes incentive shaping how students allocate academic effort. Using four years of course-level administrative records for 418 undergraduates at a Chinese public university, linked to the exact timing of their competition awards, we separate two processes with opposite distributional implications: selection, through which prior achievement raises the probability of winning and is then converted into further advantage, and compensation, through which competition involvement re-engages students whom conventional coursework serves poorly. Event-study estimates exploiting within-course variation show that selection dominates: winners already led their peers by 0.44 SD in first-year achievement before any award, and the post-award gain is roughly one-eighth as large. The returns that do exist are compensatory and concentrated in specialised coursework, and they are absent for students whose only credential is a participation prize that requires no competitive success. Because the design detects only effects of about 0.07 SD or larger, our inference rests on the selection gradient and the credential-quality contrast rather than on the magnitude of the average effect.

Keywords: academic competitions; student engagement; academic achievement; compensatory returns; credential quality; quasi-experimental design

1. Introduction

Disciplinary academic competitions have become one of the most heavily institutionalised forms of extracurricular academic activity in higher education worldwide. In China, where this study is situated, participation has grown from a marginal pursuit into a central element of undergraduate education: national, provincial, and institutional competitions now span nearly every discipline, and award records are routinely incorporated into postgraduate admission, scholarship allocation, and institutional performance metrics. Proponents argue that competitions cultivate problem-solving ability, deepen disciplinary knowledge, and foster the kind of deep engagement that classroom instruction alone rarely produces. Skeptics counter that preparation diverts time and attention from coursework, that benefits accrue disproportionately to already-advantaged students, and that awards may merely certify pre-existing ability rather than generate it. Beneath these competing claims lies a behavioural question that connects the debate to research on student engagement: how students allocate effort between coursework and competition preparation, and whether recognition changes their subsequent academic behaviour.

These competing claims map onto a deeper theoretical question that the existing literature has not adequately resolved: does competition achievement represent a signal of prior ability or a source of subsequent human capital? The distinction matters both theoretically and practically. If awards primarily signal ability, then expanding competition opportunities will do little to raise aggregate achievement and may simply re-sort students into more finely graded hierarchies. If awards generate human capital, then compe-

tition programmes constitute a genuine pedagogical investment deserving of public support. Answering this question requires observing achievement trajectories both before and after the award—precisely the design that most existing studies lack.

The empirical literature on extracurricular and cocurricular activity and academic performance is large (Cooper et al., 1999; Eccles & Barber, 1999; Guest & Schneider, 2003; Mahoney et al., 2003) but has three limitations that this study addresses. First, most studies rely on cross-sectional or two-wave survey data, comparing participants with non-participants at a single point in time. Such designs cannot distinguish selection into activities from the causal effect of activities, and the resulting estimates are vulnerable to the objection that motivated, able students self-select into both. Second, the outcome measure is typically global GPA, which aggregates across courses of widely differing difficulty, grading standards, and relevance to the activity in question. Measurement of this kind obscures the very mechanisms—knowledge transfer, time displacement—that theories emphasise. Third, the literature has been conducted almost entirely in North American and Western European contexts, where the institutional role of academic competitions differs substantially from that in East Asian systems. Research on Chinese academic competitions is growing (Fu & Li, 2025; Hou & Zhao, 2026; Zhou et al., 2025) but remains largely descriptive or relies on self-reported outcomes rather than administrative achievement records. Where effects are reported, they typically concern self-rated competencies rather than assessed performance (Ruan et al., 2025).

This study makes three contributions. First, we assemble a course-level panel linking 418 undergraduates' complete four-year course records to the exact timing of their competition awards, yielding 44,363 student-course-term observations across 248 courses. The granularity of these data permits the inclusion of course-by-term fixed effects, so that identification comes from variation within a single course offering in a single term—thereby absorbing differences in course difficulty and grading leniency that contaminate GPA-based analyses.

Second, we exploit the staggered timing of awards to implement an event-study design, distinguishing pre-award from post-award divergence in achievement. Because awards occur in different terms for different students, we can test whether award winners were already outperforming their peers before winning (consistent with signaling) or whether the gap opens only afterward (consistent with human capital formation). We complement conventional two-way fixed effects estimation with heterogeneity-robust estimators recently developed for staggered adoption designs.

Third, and most importantly, we show that the average effect is the least interesting quantity in this literature. Because selection into awards is so strong, the average return conflates two very different populations: students for whom conventional coursework is already working, and students for whom it is not. We document a compensatory gradient that reverses the cumulative-advantage expectation, and we show that credentials requiring no competitive success carry no premium at all. These findings speak directly to the policy question of whether competition programmes should be broadened, targeted, or capped, and they suggest that the answer depends on who is being asked to compete.

The two processes we examine carry opposite distributional implications. A selection process that rewards prior achievement concentrates awards among students who are already succeeding and, because awards carry formal weight in admission and funding decisions, converts that early advantage into a durable one—a cumulative process that stratifies. A compensatory process runs the other way, drawing renewed academic effort from students whom conventional coursework serves poorly and thereby equalising outcomes. Which process dominates determines whether competition policy should expand access, target lower-achieving students, or cap participation. Prior work on Chinese competition awards has mapped their expansion and stratification at the institutional level (Hou &

Zhao, 2026); we complement that account by testing the mechanism where the two processes can be most cleanly separated—at the level of individual course outcomes.

Our findings indicate that the achievement premium enjoyed by award winners is overwhelmingly a selection phenomenon. Award winners already outperform their peers by 0.44 SD before winning, and the post-award gain (0.053 SD) is roughly one-eighth as large. What modest returns exist are compensatory rather than cumulative: they accrue to students in the lowest tercile of prior achievement and to specialised coursework, and they are absent for students whose only credential is a participation prize. Given that our design has limited power to detect effects below about 0.07 SD, we build our argument on the selection gradient and the credential-quality contrast, both of which are estimated with adequate precision, rather than on the magnitude of the average effect. The remainder of the paper proceeds as follows. Section 2 develops the theoretical framework and hypotheses. Section 3 describes the institutional context. Section 4 presents the data and variable construction. Section 5 sets out the empirical strategy. Section 6 reports results, Section 7 robustness checks, and Section 8 discusses implications and limitations.

2. Theoretical Framework and Hypotheses

2.1. Human Capital versus Signaling

Two classical accounts of educational credentials offer sharply different predictions. Human capital theory treats education as productive: participation in cognitively demanding activities builds knowledge, skills, and dispositions that raise subsequent performance (Becker, 1964; Schultz, 1961). Under this view, intensive preparation for a competition—mastering advanced material, working through open-ended problems, receiving expert feedback—should generate durable gains that manifest in subsequent coursework, particularly coursework in related domains.

Signaling theory offers a rival account. Spence (1973) argued that educational credentials may serve primarily as costly-to-fake indicators of pre-existing productivity, without necessarily enhancing it; Arrow (1973) developed the closely related 'filter' account, in which education serves to reveal rather than to raise capacity. Applied to competitions, an award may certify that a student was already able, motivated, and well-prepared, conveying information to employers and admissions committees while leaving underlying capacity unchanged. Observationally, the two accounts differ in their implications for dynamics: human capital formation predicts a widening gap after the award, whereas pure signaling predicts a stable gap that predates the award.

We regard these not as mutually exclusive but as competing empirical claims whose relative weight is an open question. Our event-study design distinguishes them by examining whether the achievement differential between award winners and their classmates is already present in pre-award terms, or whether it emerges only after the award.

2.2. Student Involvement and High-Impact Practices

A second tradition emphasises the quantity and quality of student effort. Decades of research synthesised by Pascarella and Terenzini (1991, 2005) conclude that the impact of college is largely determined by the effort students devote to academic, interpersonal and extracurricular offerings. Within that tradition, Astin's theory of student involvement (1984, 1999) holds that learning and development are proportional to the physical and psychological energy students devote to their education. Carini et al. (2006) provide direct evidence that engagement measures predict performance on standardised outcomes, though the estimated associations are typically modest. Kuh (2008) extended this logic by identifying high-impact practices—activities demanding substantial time and effort, in-

volving interaction with faculty and peers, and providing frequent feedback—as particularly potent. Academic competitions exhibit most of these features: sustained engagement over months, close mentoring by faculty, collaborative problem-solving, and iterative feedback on performance.

This literature also suggests a boundary condition. Reviews of extracurricular involvement consistently find that moderate engagement is compatible with or beneficial to academic performance, whereas excessive engagement generates time conflicts and stress that undermine it (Fredricks & Eccles, 2006, 2008; Marsh & Kleitman, 2002). The relationship is plausibly non-monotonic, which motivates our examination of award intensity.

2.3. The Zero-Sum and Displacement Perspective

A third perspective emphasises opportunity cost. Coleman (1961) argued that student time is finite and that investment in non-academic pursuits necessarily competes with academic work—a zero-sum logic. In higher education, this implies that competition preparation during a given term should reduce the time available for coursework in that term, producing a short-run depression of grades even if longer-run benefits materialise. This perspective generates a distinctive dynamic prediction: a dip in the term of preparation followed by recovery or overshooting thereafter.

2.4. Hypotheses

Synthesizing these frameworks, we advance the following hypotheses:

H1a (Human capital). If competition awards generate skills, award winners' course achievement will improve significantly in post-award terms relative to pre-award terms and relative to non-winners.

H1b (Signaling). If awards primarily certify prior ability, award winners will already outperform their classmates in pre-award terms, and the differential will not widen after the award.

H2 (Dynamic trade-off). If displacement operates in the short run and skill formation in the long run, achievement will decline in the term of preparation and rise subsequently, yielding a J-shaped response path.

H3 (Domain specificity). If knowledge transfer underlies any effect, gains will be larger in courses within the competition's disciplinary domain than in unrelated coursework.

H4a (Cumulative advantage). If competitions function as enrichment for those already succeeding, gains will be concentrated among students in the highest tercile of prior achievement, implying that competitions reinforce existing hierarchies.

H4b (Compensatory). If competitions function as an alternative engagement pathway, gains will be concentrated among students in the lowest tercile of prior achievement, for whom conventional coursework is least effective. This would support broadening access rather than targeting the already successful.

H5 (Credential quality). If returns reflect involvement rather than mere certification, credentials that require no competitive success, awarded for participation alone, should carry no achievement premium, whereas competitively earned awards should.

We specify H4a and H4b as explicit competitors because the distributional question carries the clearest policy content. We regard H5 as central to interpretation: it provides a direct test of whether awards function as signals, since under a pure signalling account any credential should certify pre-existing ability, whereas under an involvement account only competitively earned awards should carry a premium.

3. Institutional Context

Understanding the Chinese case requires brief institutional background, as the role of academic competitions differs markedly from that in Western systems.

At the apex of the system sit a small number of national flagship competitions, of which two dominate institutional attention and resource allocation: the China International College Students' Innovation Competition and the "Challenge Cup". The former is organised by the Ministry of Education together with other central government bodies; the latter is organised jointly by the Communist Youth League Central Committee, the China Association for Science and Technology, the Ministry of Education, and the All-China Students' Federation. Both operate through a hierarchical structure in which institutional selection precedes provincial and national rounds, so that a national award normally implies success at two earlier stages and a preparation period extending well beyond a single term. This multi-tier design matters for our purposes because it generates substantial variation in how much effort a given award represents.

Beneath these flagship events lies a much denser layer of discipline-specific competitions, and it is this layer rather than the flagship events that generates most of the awards observed in our data. In food science and related fields these include product-development and innovation contests organised by national engineering research centres and industry associations, food-science knowledge competitions, and national contests in life sciences and in engineering practice training. Such events are typically organised by national professional societies, by teaching guidance committees operating under the Ministry of Education, or by industry bodies, rather than by government directly, and they vary considerably in scale, selectivity and prestige. That heterogeneity is central to our analysis: the same administrative register records both highly selective national contests and events in which enrolment alone yields a certificate.

The incentive structure is what most clearly distinguishes this setting from systems in which competitions are voluntary enrichment. Awards carry formal weight in two decisions of direct consequence to undergraduates. The first is admission to postgraduate study by recommendation, which exempts selected students from the national entrance examination and is allocated substantially on the basis of a composite score of which competition awards are one component. The second is the annual scholarship and merit ranking, in which award points similarly enter a composite index. Students therefore face an institutionalised return to competing that is largely absent in systems where awards carry no formal weight. The scale of participation in our setting is consistent with this: 35.4 % of the cohort received at least one recorded award during their studies.

Two features of this context bear on interpretation. First, because awards carry formal weight in postgraduate admission and scholarship decisions, the incentive to participate is stronger than in systems where competitions are purely voluntary enrichment. This raises the stakes of the selection problem and makes pre-award trend testing especially important. Second, the institutionalisation of competitions means that participation is frequently organised through faculty mentoring and training programmes, which may themselves deliver instructional content—a mechanism through which human capital formation would operate.

4. Data and Variables

4.1. Data Source and Sample

We combine two administrative data sources from a public university in southern China. The first is the university's complete undergraduate course-grade registry, which contains every course taken by every student in the 2021-entry cohort of the College of Food Science and Technology, together with the course code, term of enrollment, raw score, credit value, and course category. The second is the registry of competition awards

maintained by the Academic Affairs Office, which records the student identifier, the term in which the award was received, the competition name, and the award level. All student identifiers were replaced with random codes prior to analysis.

The analytical sample is constructed as follows. We begin with 47,636 raw grade records covering 418 students in two majors (Food Science and Engineering, 210 students; Food Quality and Safety, 208 students) across 14 class cohorts. We exclude 482 invalid records (absent, deferred, disqualified, cheating), convert 1,958 letter-graded records using the institutional grade-point equivalence (excellent = 95, good = 85, satisfactory = 75, pass = 65), drop 28 zero-score records that almost certainly denote non-submission, restrict to the eight regular terms (excluding 694 records from other academic years), retain only first attempts for retaken courses (excluding 731 records), and drop 1,338 observations in course-term cells with fewer than ten enrollees. Appendix A reports the full sequence. The resulting sample comprises 418 students, 248 courses and 44,363 student-course-term observations, of whom 148 (35.4 %) received at least one competition award. Retention is 93.1 % of the original records.

4.2. Dependent Variable: Course-Level Standardised Achievement

Raw scores are not comparable across courses: some courses are graded leniently and others harshly, and these differences would otherwise be confounded with student ability. We therefore standardise achievement within course offerings. Specifically, for student i in course c during term t, we compute:

$$Z_{ict} = \frac{Score_{ict} - Mean_{ct}}{SD_{ct}} \quad (1)$$

where the mean and standard deviation are taken over all students enrolled in the same course in the same term. The resulting measure expresses a student's position relative to classmates in that particular offering, and is comparable across courses and terms. Where a course-term cell contains fewer than ten students, we compute the standardisation at the course level instead; 1.7 % of observations use this fallback, and 2.9 % (predominantly online elective courses taken by a single student) are dropped. Because a small number of courses contain extreme low scores, the standardised measure is winsorised at plus or minus four standard deviations; this affects 110 observations (0.25 %). As a further robustness check we also employ within-course percentile ranks, which are insensitive to outlying grades.

4.3. Treatment Definition and Timing

Our treatment is the receipt of a competition award, and the timing variable τ_i is the term of a student's first award. Students who never receive an award form the comparison group. Constructing this variable required two decisions that materially shape interpretation.

The first concerns award quality. Inspection of the award registry reveals that a substantial share of recorded "awards" carry no competitive content. Most prominent is the National Academic English Vocabulary Competition (NAEV), in which participation alone yields a "Winning Prize"; all 46 participation-award records in our data come from this competition, and 43 of them fall in a single term. Treating such credentials as achievements would equate winning with signing up. We therefore exclude participation awards (identified by the phrases Participated, Winning Prize, 优胜奖, and related markers), which removes 19 students who hold no other award and leaves a main treatment group of 129 students (30.9 % of the cohort). We retain these students in a separate falsification exercise reported in Section 7.

The second concerns sample period. Our award records span academic years 2020–2021 through 2024–2025, which together cover all eight terms of the cohort's programme.

We nevertheless restrict the main analysis to terms one through six, for a reason established by the data rather than by availability. The 2024–2025 registry records only twelve awards to members of this cohort, all to students who had already won in an earlier term; fourth-year students had, in effect, stopped competing. The binding constraint is instead the composition of the final year. In term eight, students enrol in an average of just 2.2 courses, overwhelmingly graduation theses, internships, and political education, graded on a fundamentally different logic from taught coursework, and in term seven the average is 9.3 courses with a similarly atypical mix. Including these terms therefore adds little information about academic performance while introducing substantial measurement heterogeneity. Estimates from the full eight-term panel are reported as a robustness check.

Because our administrative records document awards but not participation, students who competed without winning are subsumed in the comparison group. These students are likely more similar to winners in motivation and effort than are non-participants, so our estimates should be interpreted as a conservative lower bound on the effect of competition involvement.

Awards are staggered: students win in different terms, with first awards distributed across term 1 (13 students), term 2 (34), term 3 (17), term 4 (30), term 5 (29), and term 6 (6). This variation identifies our event-study estimates.

4.4. Control Variables

Because we include student fixed effects, all time-invariant characteristics—prior achievement at entry, family background, gender, region of origin, and stable personality traits—are absorbed and cannot bias our estimates; indeed they cannot be separately identified. This is a strength of the within-student design rather than a limitation. We additionally control for the number of courses taken in each term, which captures variation in academic load and serves as an observable proxy for the time demands facing the student.

To proxy pre-treatment ability for purposes of heterogeneity analysis and matching, we construct each student's mean standardised achievement during the first academic year (terms 1–2). Because this measure precedes any award in our data, it is pre-determined with respect to treatment, and we use it to partition students into terciles of prior academic standing. Our records also include gender, major, and class cohort. These are time-invariant and therefore absorbed by student fixed effects in the pooled models; we nonetheless use them to estimate separate regressions by subgroup, which permits effect heterogeneity along these dimensions even though their main effects are not separately identified.

4.5. Descriptive Statistics

Before turning to the estimates, we describe the treatment group. Figure 1 summarises its construction and composition. Panel (a) plots the distribution of first-award timing across terms one through six; awards are staggered throughout the period rather than clustered in a single term, which is what makes the event-study design feasible. Panel (b) reports group sizes: 129 award winners against 289 non-winners. Panel (c) decomposes awards by type, showing that innovation and STEM competitions and academic knowledge contests together account for the largest share, with sports and cultural activities also well represented. This composition matters for interpretation, because it indicates that the awards in our data span a wide range of activities rather than a single domain.

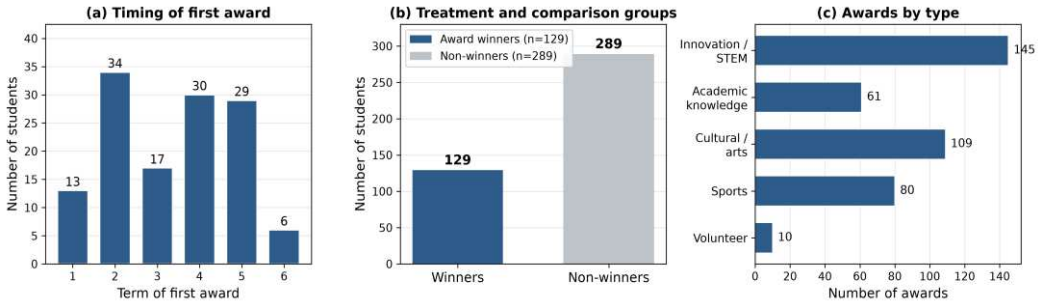


Figure 1. Construction and composition of the treatment group. (a) Distribution of the term of first award. (b) Size of the treatment and comparison groups. (c) Number of awards by competition type.

Table 1 reports descriptive statistics and compares award winners with non-winners on pre-treatment characteristics.

Table 1. Descriptive statistics and pre-treatment differences between award winners and non-winners.

Variable	Full sample	Award winners	Non-winners	Difference
Standardised achievement (Z)	0.004 (0.976)	0.318 (0.841)	-0.142 (0.999)	0.460
Raw score	83.84 (9.86)	86.22 (8.48)	82.74 (10.26)	3.483
First-year mean Z (pre-treatment)	-0.012 (0.506)	0.291 (0.375)	-0.152 (0.497)	0.442
Courses taken per term	16.16 (2.76)	16.33 (2.56)	16.09 (2.85)	0.240
Credits per term	23.89 (5.48)	24.07 (5.37)	23.81 (5.54)	0.260
Share of specialized courses	0.383 (0.486)	0.383 (0.486)	0.384 (0.486)	-0.001
Number of students	418	129	289	-
Student-course-term observations	39,633	12,573	27,060	-

Note: Standard deviations in parentheses. Sample restricted to terms 1–6. Award winners are students with at least one competitively earned award; participation-only credentials are excluded. Final column reports raw mean differences. Descriptive counts refer to the full terms 1–6 sample; the estimation samples behind Tables 2–4 are marginally smaller because singleton course × term cells are dropped in estimation.

Table 1 reveals pronounced positive selection into awards. Students who go on to win score 0.318 SD above the cohort mean in standardised achievement, against -0.142 SD for those who do not—a gap of 0.46 SD, or roughly 3.5 percentage points in raw scores. Crucially, this gap is already present before any award: on first-year achievement, which precedes every award in our data, winners lead non-winners by 0.44 SD (Figure 2). Winning is therefore strongly predicted by prior academic standing, and any comparison failing to account for this would conflate the consequences of awards with the characteristics of those who receive them. On observable time-varying dimensions the two groups are more similar: winners take 16.3 courses and 24.1 credits per term against 16.1 and 23.8 for non-winners, and enrol in specialised courses at virtually identical rates. These small differences are reassuring but do little to address selection on unobservables.

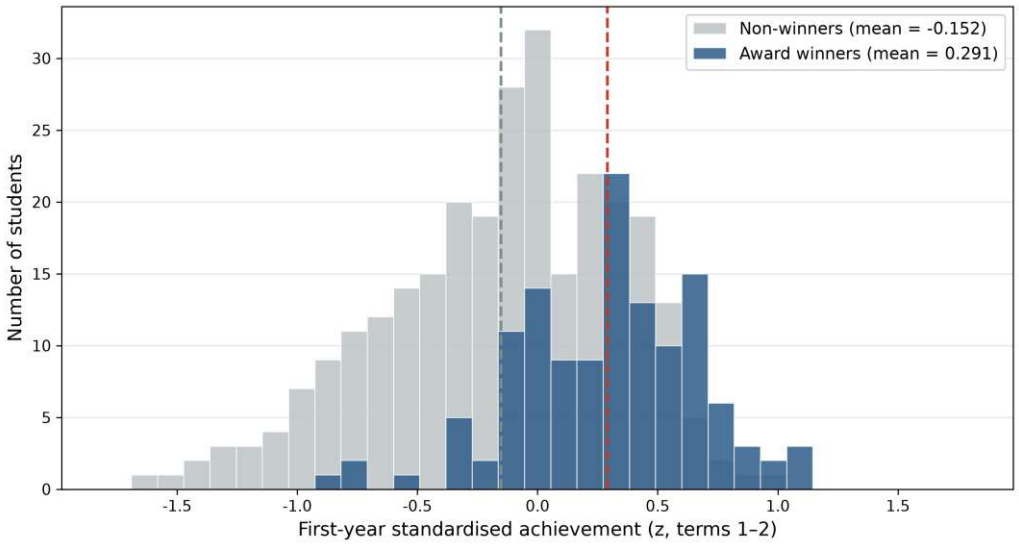


Figure 2. Positive selection into awards. Distribution of first-year standardised achievement (terms 1–2), measured before any award was received, for students who go on to win an award and those who do not. Vertical lines mark group means; the gap of 0.44 SD is the selection gradient referred to throughout.

The patterns in Table 1 motivate our empirical strategy directly: because the achievement gap predates treatment, identification must come from within-student variation over time rather than from cross-sectional comparison.

5. Empirical Strategy

5.1. Course-Level Event Study

Our baseline specification is an event study estimated on the student-course-term panel:

$$Z_{ict} = \alpha_i + \gamma_{ct} + \Sigma_k \beta_k \cdot 1[t - \tau_i = k] + \theta \cdot X_{it} + \varepsilon_{ict} \quad (2)$$

where α_i denotes student fixed effects, γ_ct denotes course-by-term fixed effects, and 1[t − τ_i = k] indicates that term t falls k periods relative to student i's first award. The term immediately preceding the award (k = −1) is the reference period and its coefficient is normalised to zero. X_it is the vector of time-varying controls described above. Standard errors are clustered at the student level, following the guidance of Bertrand et al. (2004) on inference in difference-in-differences designs with serially correlated outcomes.

The course-by-term fixed effects are consequential. They restrict comparisons to students enrolled in the same course in the same term, so that identification derives from within-classroom variation rather than from course selection. Differences in course difficulty, instructor grading standards, and term-specific shocks—including any disruption common to all students in a given term—are thereby absorbed.

The coefficients β_k for k < 0 serve as placebo tests: under the identifying assumption that, absent the award, winners and non-winners would have followed parallel trajectories, these coefficients should be statistically indistinguishable from zero. Coefficients for k ≥ 0 trace the dynamic treatment effect. We bin the endpoints of the event window to avoid imprecise estimates from sparsely populated extreme periods.

5.2. Heterogeneity-Robust Estimation

Recent econometric research demonstrates that conventional two-way fixed effects regression can produce biased estimates in staggered adoption designs when treatment

effects vary across cohorts or over time, because already-treated units implicitly serve as controls for later-treated units (de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021). We therefore complement the baseline with the interaction-weighted estimator of Sun and Abraham (2021), which saturates the event study with cohort-by-relative-time effects and aggregates them using cohort shares, so that already-awarded cohorts never serve as controls for later ones; related group-time estimators (Callaway & Sant'Anna, 2021) address the same concern. Results from this robustness exercise are reported in Section 7. Our choices follow the synthesis of recent practice in Roth et al. (2023).

The interaction-weighted estimator requires a balanced panel aggregated to the student-term level, so we average standardised achievement within student-term cells for this portion of the analysis. The aggregation necessarily sacrifices the course-by-term fixed effects, so we regard the two approaches as complementary: the course-level specification offers finer control for unobserved course heterogeneity, while the heterogeneity-robust estimators protect against negative weighting under effect heterogeneity.

5.3. Threats to Identification

Three threats deserve explicit discussion. The first is time-invariant selection: students who win awards may differ systematically from those who do not. Student fixed effects fully address this concern, as noted above.

The second is time-varying selection: a student may experience a contemporaneous motivational or effort shock that raises both the likelihood of winning and course performance. We address this through the pre-trend tests embedded in the event study, through a placebo exercise that artificially advances award timing, and through sensitivity analysis described below.

The third is anticipation: students typically prepare for competitions before winning, so the no-anticipation assumption underlying difference-in-differences may be violated. We probe this by examining the coefficient at $k = -1$ and by re-estimating the model with $k = -2$ as the reference period. If preparation effects are present, our post-award estimates should be understood as capturing the combined effect of preparation and award rather than the award per se—an interpretation we adopt explicitly in the discussion.

5.4. Sensitivity Analysis

We also note that the parallel-trends assumption itself can be relaxed rather than assumed. Recent work develops estimators that deliver valid inference under bounded violations of parallel trends, allowing researchers to report how conclusions change as the permitted violation grows (Rambachan & Roth, 2023). Applying such methods to our setting would be a natural extension, though we leave it to future work.

Because our administrative data lack direct measures of ability and motivation, we assess the robustness of our estimates to unobserved selection using the bounding procedure of Oster (2019). This approach calculates the degree of selection on unobservables, relative to selection on observables, that would be required to explain away the estimated effect. We report the implied δ statistic, with values exceeding one conventionally interpreted as indicating robustness.

We additionally conduct placebo tests that reassign award timing to two and three terms earlier than the true date; under the null of no pre-award divergence, these exercises should yield null results. Where the number of treated units is modest, we supplement conventional cluster-robust inference with wild cluster bootstrap procedures, which provide more reliable inference when the number of clusters is small.

5.5. Statistical Power and Interpretive Caution

We report power calculations because they shape how our results should be read. With 418 students and clustering at the student level, the minimum effect detectable at conventional levels is approximately 1.96 times the standard error. For our preferred specification this threshold is 0.050 SD, whereas achieving 80 % power would require an effect of roughly 0.072 SD. Our estimated average effect of 0.053 SD sits just above the significance threshold but well below the level at which the design offers reliable detection. Evaluated ex post, the power of our design against an effect of this magnitude is 0.55, implying that a replication drawing a different cohort would fail to detect a genuinely present effect of this size nearly half the time.

Two consequences follow, and we state them plainly. First, our point estimates should be interpreted with corresponding caution: where an underpowered design does yield significant results, the estimated magnitude tends to overstate the true effect, because only estimates large enough to clear the threshold are detected. Second, the same applies to the subgroup estimates. The compensatory gradient for the lowest achievement tercile (0.217, SE = 0.112) is only marginally significant and has ex post power of 0.49, so we treat its magnitude as suggestive rather than established; reaching 80 % power for the average effect we estimate would require approximately 760 students, close to twice our sample. This is the central reason we do not build our argument on the magnitude of the average effect. The magnitude itself is also small by the standards of the field: recent recalibration of effect-size benchmarks for education interventions places 0.05 SD in the small-but-meaningful range and 0.20 SD at the upper end of what field studies typically detect (Kraft, 2020).

We therefore organise our interpretation around the quantities our design can estimate precisely. The selection gradient documented in Table 1, a 0.44 SD advantage in first-year achievement prior to any award, is a descriptive comparison measured on the full cohort and is not subject to these power concerns. The contrast between competitively earned awards and participation credentials is likewise informative regardless of the precision of the average effect, since the participation-award estimate has an upper bound of roughly 0.07 SD, essentially the detection limit of this design. Accordingly, we do not rest our conclusions on the magnitude of the average post-award effect, and we say so explicitly rather than presenting it as the headline finding.

6. Results

6.1. Main Results

Table 2 reports our main estimates. Column (1) presents the static difference-in-differences estimate pooling all post-award periods on the full eight-term panel; column (2) restricts to terms one through six, our preferred sample; column (3) adds time-varying controls; and column (4) restricts the comparison group to students who never received any award.

Table 2. Difference-in-differences estimates of the post-award effect of a competition award on course achievement.

	(1) Full panel	**(2) Terms 1–6**	**(3) + controls**	**(4) Never-awarded control**
Post-award effect	0.040	0.051**	0.053**	0.051*
	(0.0250)	(0.0255)	(0.0256)	(0.0258)
Student fixed effects	Yes	Yes	Yes	Yes

	(1) Full panel	(2) Terms 1–6	(3) + controls	(4) Never-awarded control
Course × term fixed effects	Yes	Yes	Yes	Yes
Time-varying controls	Yes	No	Yes	Yes
Observations	44,317	39,597	39,597	37,725
Treated students	129	129	129	129

Note: Dependent variable is within-course-term standardised achievement, winsorised at ±4 SD. Standard errors clustered at the student level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find a small but statistically detectable positive association between winning a competition and subsequent course achievement. In our preferred specification (column 3), award winners score 0.053 standard deviations higher in post-award terms than in pre-award terms, relative to the change among non-winners (SE = 0.026, $p = 0.038$). To convey magnitude, this corresponds to approximately 0.4 percentage points in raw scores, or roughly one-twentieth of the within-course standard deviation of grades. The effect is modest by any standard.

The attenuation in column (1) is informative. When terms seven and eight are included, the estimate falls to 0.040 and is no longer statistically significant ($p = 0.115$). We do not attribute this primarily to missing award records, since the separately obtained 2024–2025 registry shows that final-year awards were rare in this cohort and all went to students who had already won earlier. The more consequential difference is compositional: including terms dominated by thesis supervision, internships, and political education introduces measurement heterogeneity that dilutes the signal, a pattern consistent with the attenuation we observe. We regard the terms 1–6 specification as the more credible one and report the full-panel estimate as a robustness check. We note immediately, and return to this in Section 5.5, that an effect of this magnitude is close to the minimum our design can reliably detect, so the point estimate should be read as indicating the presence and direction of a small association rather than as a precise measurement of its size. Column (4), which restricts the comparison group to students who never received any award, yields a very similar estimate of 0.051 (SE = 0.026, $p = 0.050$), which is reassuring given that this specification discards the not-yet-awarded comparison group entirely.

6.2. Event-Study Estimates and Parallel Trends

Figure 3 presents the unadjusted mean achievement of award winners by event time, and Figure 4 plots the corresponding event-study coefficients with 95 % confidence intervals, with the term preceding the award normalised to zero.

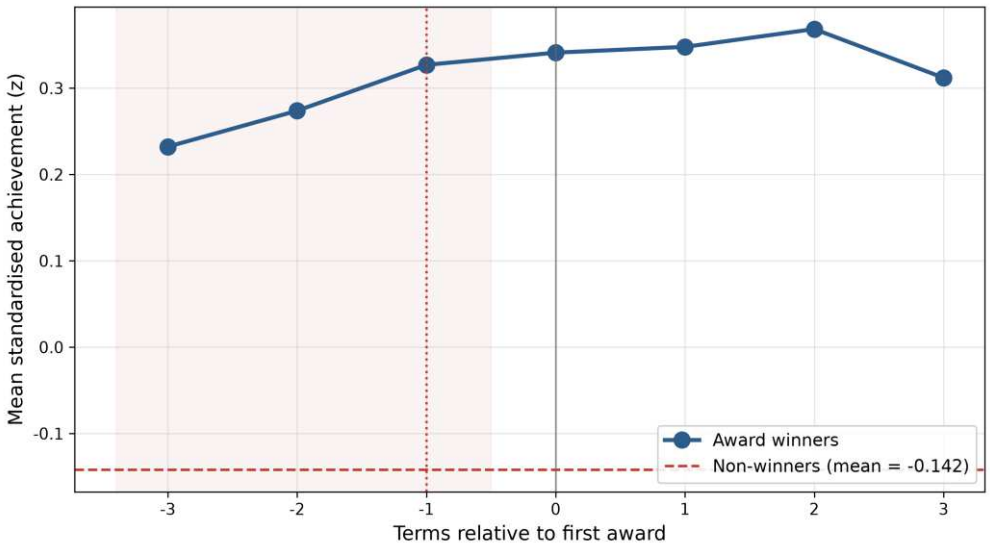


Figure 3. Unadjusted mean standardised achievement of award winners by event time. The horizontal dashed line is the mean for non-winners. Unlike Figure 4, this figure is not regression-adjusted and does not remove course-by-term composition; it is provided as a descriptive complement showing the dip-and-recovery pattern in the raw means.

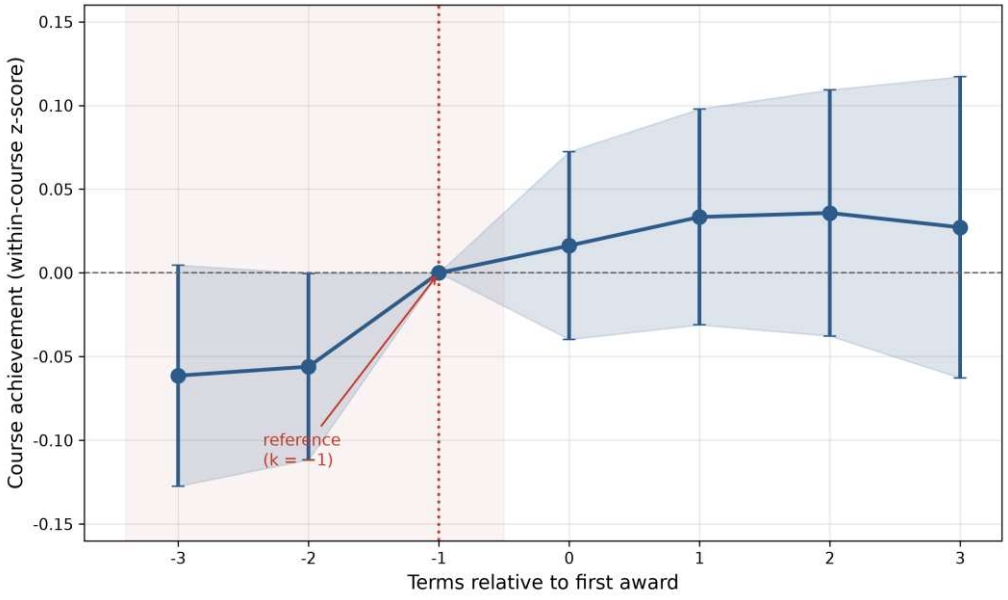


Figure 4. Event-study estimates of the effect of a competition award on course achievement. Points are coefficients from the specification in column (3) of Table 2; bars are 95% confidence intervals based on standard errors clustered by student. The term preceding the award (k = −1) is the reference period. The shaded area marks pre-award periods.

The dynamic pattern is not the one our hypotheses anticipated, and we report it candidly. Post-award coefficients are uniformly positive but individually insignificant, ranging from 0.016 to 0.036. More importantly, the two available pre-award coefficients are negative and of similar magnitude (−0.056 at k = −3 and −0.054 at k = −2, both marginally significant). A joint test of the pre-award coefficients yields Wald = 5.20 ($p = 0.075$), which does not reject parallel trends at the 5 % level, though the marginal significance warrants the sensitivity analysis reported below.

Table 3. Event-study estimates of the effect of a competition award on course achievement.

Event time k	Coefficient	Std. error	p-value	Event time k	Coefficient	Std. error	p-value
−3	−0.0614	0.0337	0.069	0	0.0163	0.0286	0.568
−2	−0.0560	0.0284	0.049	+1	0.0334	0.0329	0.311
−1 (ref.)	0.000	—	—	+2	0.0358	0.0375	0.340
				+3	0.0272	0.0459	0.554

Note: Estimates from the specification in column (3) of Table 2. Joint test of pre-award coefficients ($k < 0$): Wald = 5.20, df = 2, p = 0.075.

Two readings of this pattern are available. Note first that the negative pre-award coefficients contrast with the positive level difference documented in Table 1: winners are better students on average, yet within-student they appear to dip in the terms immediately preceding an award. Under the first reading, this dip reflects preparation costs, as students divert effort toward competition preparation two to three terms before winning, and the post-award coefficients capture recovery rather than a genuine gain. Under the second, students experiencing a dip in academic performance respond by redirecting effort toward extracurricular competition—a compensatory investment pattern—and subsequently recover. Our data cannot definitively distinguish these accounts. What we can say is that the post-award estimates should not be interpreted as clean causal effects; we therefore frame our conclusions in Section 8 with appropriate restraint and quantify the sensitivity of our estimates to pre-trend violations.

6.3. Heterogeneity

Table 4 examines heterogeneity along four dimensions. Although the average effect is small, the pattern across subgroups is systematic and theoretically informative.

Three results stand out. First, the effect is concentrated in specialised coursework: award winners score 0.145 standard deviations higher in specialised courses after winning (p = 0.045), while estimates for general education, ideological, and practical coursework are smaller and statistically indistinguishable from zero. This is consistent with domain-specific knowledge transfer, the mechanism central to the human capital account, although the absence of a comparable effect in basic specialised courses ($\beta = -0.073$) complicates the picture.

Second, and most striking, effects are largest for students in the lowest tercile of prior achievement. Award winners in the bottom tercile gain 0.217 standard deviations (p = 0.054), roughly four times the estimate for the top tercile (0.052, p = 0.072). This compensatory pattern runs against the cumulative-advantage expectation derived from expectation-states theory and instead suggests that competition involvement may function as an alternative route to academic engagement for students who are not thriving in conventional coursework. We note that the estimate for the low-achieving group is imprecisely measured (SE = 0.112) and, as Section 5.5 details, marginally powered; we therefore treat the four-fold ratio as suggestive of a compensatory pattern rather than as a precise measurement of its magnitude.

Third, the effect is concentrated among women ($\beta = 0.071$, p = 0.012) and absent among men ($\beta = 0.020$, p = 0.717). We are reluctant to over-interpret this difference, as it

may reflect gender differences in the types of competitions entered rather than differential returns; our data do not permit a clean separation of these channels.

Table 4. Heterogeneous post-award effects by course type, prior achievement tercile, gender, and award level.

Panel	Coefficient	Std. error	p-value	N
A. By course type				
Specialized courses	0.1453**	0.0723	0.045	7,954
Basic specialized	−0.0727	0.0571	0.204	7,223
General education	0.0689	0.0546	0.208	8,807
Ideological/political	0.0670	0.0497	0.179	5,230
Practice and laboratory	0.0484	0.0381	0.204	10,370
B. By prior achievement tercile				
Lowest tercile	0.2165*	0.1115	0.054	13,081
Middle tercile	0.0765	0.0520	0.143	12,897
Highest tercile	0.0515*	0.0284	0.072	12,988
C. By gender				
Female	0.0713**	0.0280	0.012	22,301
Male	0.0203	0.0559	0.717	17,246
D. By award level				
Institution-level or below	0.0785*	0.0458	0.091	6,672
Provincial-level	0.0487	0.0535	0.368	3,952
National-level	−0.0505	0.0663	0.456	1,781

Note: Each cell reports a separate estimate of the post-award effect within the indicated subsample, with student and course-by-term fixed effects. Standard errors clustered at the student level. * $p < 0.10$, ** $p < 0.05$.

6.4. Mechanisms

We probe two mechanisms. To test the time-displacement hypothesis, we re-estimate our model with the number of courses taken in each term as the outcome. If competition preparation crowds out study time, award winners should reduce their course load in the award term. The estimate is negative but insignificant ($\beta = -0.094$ courses, SE = 0.104, $p = 0.367$), offering at most weak support for displacement. The point estimate implies that winners take roughly one-tenth of a course fewer per term, a magnitude too small to account for meaningful grade changes.

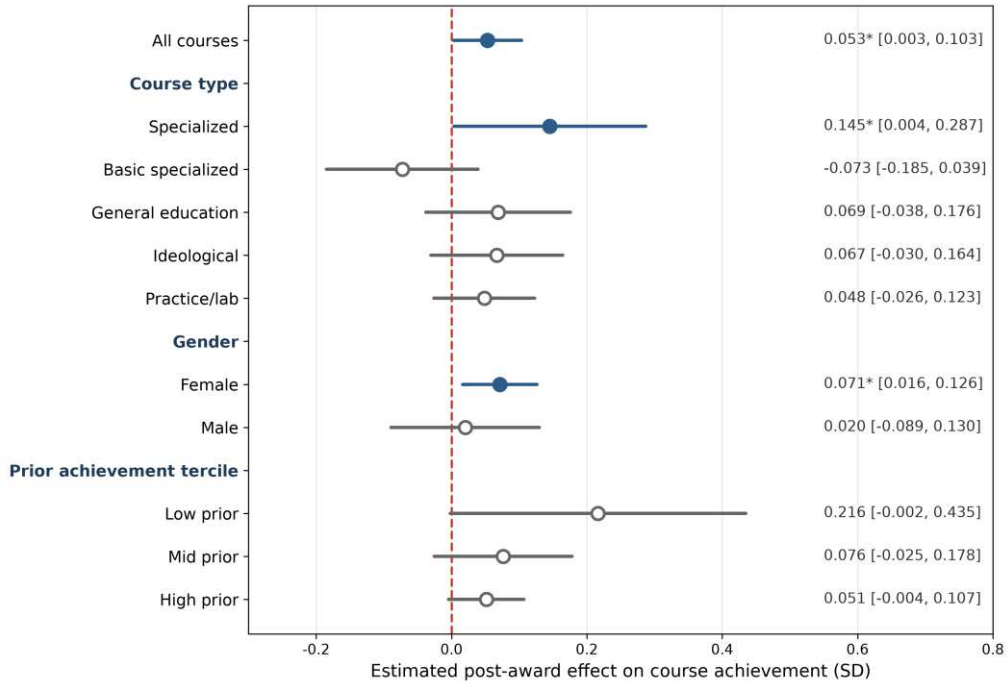


Figure 5. Heterogeneous effects by course type, prior achievement, and gender. Each point is a separate estimate of the post-award effect within the indicated subsample; bars are 95% confidence intervals. Solid points are significant at the 5% level; hollow points are not. Sample restricted to terms 1–6.

To test the knowledge-transfer hypothesis, we compare effects across course types as reported in Panel A of Table 4. The concentration of gains in specialised courses is consistent with transfer, though we emphasise that this is indirect evidence: our administrative data contain no measure of study time, motivation, or the specific competencies competitions are intended to build, and we cannot rule out that specialised courses are simply more responsive to any change in student effort.

7. Robustness Checks

We conduct seven exercises. First, we re-estimate the model using the full eight-term panel rather than terms one through six; the estimate attenuates from 0.053 to 0.040 and is no longer statistically significant ($p = 0.115$). The explanation for this attenuation is compositional and is set out in Section 6.1. Second, we cluster standard errors two ways, by student and by course; inference is unchanged. Third, we drop the winsorisation and use untrimmed standardised scores; results are substantively identical. Fourth, we restrict the comparison group to students who never received any award, excluding not-yet-awarded students entirely; the estimate is 0.051 (SE = 0.026, $p = 0.050$), essentially unchanged.

Fifth, we conduct placebo tests that reassign award timing to two and three terms earlier than the true date. These yield estimates of 0.050 ($p = 0.112$) and −0.030 ($p = 0.453$), neither significant at conventional levels. Sixth, and most important, we exploit the distinction between genuine awards and participation credentials. Re-estimating the model on the 19 students whose only recorded "award" is a participation credential—an NAEV certificate awarded for signing up—yields an estimate of −0.055 (SE = 0.066, $p = 0.405$). The contrast is instructive: credentials that require no competitive success carry no academic premium, while those that do are associated with modest gains. This pattern argues against a pure signalling interpretation, under which any credential would certify pre-existing ability, and is more consistent with the involvement account.

Seventh, we re-estimate the event study using the interaction-weighted estimator of Sun and Abraham (2021) on the student-term panel. The implied average post-award effect over the available post-award terms is 0.018 (SE = 0.032) and is statistically indistinguishable from zero. This estimate is smaller than the two-way fixed effects baseline, which is what one expects if conventional weighting places disproportionate weight on the cohorts with the largest pre-award declines; it reinforces, rather than undermines, our decision to rest the argument on the selection gradient and the credential-quality contrast.

A final caveat concerns robustness to unobserved selection. Because our estimates are small and the pre-trend test is marginal, we do not place weight on the Oster bounding exercise; the implied δ statistic is 0.049, effectively zero, indicating that the estimate is fragile to selection on unobservables. We regard the honest conclusion to be that the average effect is small and not robustly identified, and we turn in Section 8 to the subgroup patterns, which are more consistent across specifications.

8. Discussion

Our study set out to ask whether competition awards represent a signal of prior ability or a source of subsequent human capital. The honest answer is that our data support neither account strongly, and we think this near-null result is itself the contribution. More precisely, our data allow us to say what awards are not: they are not primarily credentials that certify an advantage their holders already enjoyed before receiving any award, and they are not powerful engines of human capital formation either. What remains is a small, unevenly distributed return that accrues disproportionately to the students conventional coursework serves least well.

On the signalling account, award winners should already outperform their classmates before winning. In levels, they clearly do: as Table 1 shows, winners lead non-winners by 0.44 SD on first-year achievement, before any award is received. Yet the dynamic evidence complicates this reading. Within students, performance dips in the two to three terms preceding an award before recovering afterward, a pattern more consistent with preparation costs than with pure certification. On the human capital account, awards should generate durable gains that widen the gap. We find post-award gains that are positive but small—about 0.05 standard deviations, roughly one-twentieth of the within-student dispersion in course grades—and individually insignificant at each horizon. A dispassionate reading is that winning a competition is associated with a marginal improvement in course performance, and that most of the achievement premium enjoyed by award winners predates the award rather than results from it. This conclusion closely parallels the one reached by Guo and Zhang (2023), who combine national survey data with award records from ten high-level national competitions and report that high-level competitions select out students with excellent academic performance but that winning brings little additional gain. Our design differs from theirs in using course-level administrative records and exploiting the timing of awards, but the convergence is reassuring: two very different data structures and identification strategies point to the same conclusion.

Three findings nonetheless carry substantive weight. First, effects are concentrated in specialised coursework ($\beta = 0.145$, $p = 0.045$) and are absent in general education and practical courses, a pattern consistent with domain-specific knowledge transfer. Second, and contrary to the cumulative-advantage expectation, gains are largest for students in the lowest tercile of prior achievement ($\beta = 0.217$, $p = 0.054$) and smallest for the highest tercile. Third, the distinction between genuine awards and participation credentials matters: students whose only credential is a participation prize show no gain at all ($\beta = -0.055$, $p = 0.405$), whereas students with competitive awards show modest positive gains.

Taken together, these patterns suggest a more nuanced picture than the signal-versus-human-capital dichotomy allows. Competition awards appear to function less as credentials that certify ability and less as interventions that build it wholesale, and more as one among several routes through which students who are not thriving in conventional coursework can re-engage. The compensatory gradient is the most policy-relevant finding we have: if competition involvement benefits lower-achieving students more than their higher-achieving peers, then broadening access may have distributional value even where average returns are modest. It also sits naturally alongside recent longitudinal evidence that engagement predicts Chinese undergraduates' learning outcomes through motivational rather than purely behavioural channels (Yang et al., 2026), which would help explain why the students with the most to gain are those least engaged by conventional coursework. A plausible mechanism, though one our data cannot isolate, is that recognition strengthens academic self-efficacy for students whose prior record gives them little occasion for confidence, raising the effort they subsequently devote to coursework; testing this interpretation directly would require measures of motivation and self-efficacy that administrative records do not contain.

These findings contribute to the literature in three ways. Conceptually, we show that the debate over whether awards signal or produce ability has been framed too narrowly. In our setting the more consequential question is distributional: returns are compensatory, not cumulative. This reframing matters because the policy conclusions differ sharply; a programme justified on aggregate returns deserves broad expansion, whereas one justified on compensatory returns deserves targeted outreach. Methodologically, we demonstrate the value of course-level administrative data and of distinguishing genuine awards from participation credentials—a distinction that, to our knowledge, has not been drawn in prior work and that materially changes the estimate. Substantively, we provide course-level evidence from a Chinese institution, where competitions are institutionally consequential but where evidence has been largely descriptive.

For practice, our results counsel restraint. Competition programmes should not be justified on the expectation that they raise course grades: the average return is small, and for national-level award winners the point estimate is negative ($\beta = -0.051$, $p = 0.456$), perhaps because the most intensive competitors divert substantial time from coursework. What our results do support is a targeted rather than universal rationale. Because gains concentrate among lower-achieving students and in specialised coursework, competition programmes may be most valuable as an alternative engagement pathway for students underserved by conventional instruction, rather than as enrichment for those already succeeding.

8.1. Limitations

Several limitations qualify our conclusions. First, our data come from a single institution. The university is provincially administered rather than a national 'Double First-Class' institution, and its food science programmes rank in the 100–150 band in national discipline rankings. This positioning cuts both ways for external validity. On the one hand, it makes the setting arguably more representative of Chinese undergraduate education than the elite institutions that dominate the literature, since the large majority of Chinese undergraduates study at institutions of this tier. On the other hand, food science and technology is an applied, laboratory-intensive discipline with a distinctive course structure—a large practical and experimental component, close industry linkage, and competition activity organised around product development rather than abstract problem solving. Returns to competition involvement in theoretical disciplines, or in the humanities and social sciences, may follow a different pattern, and our estimates should not be extrapolated to those settings without direct evidence. Second, we observe awards but not

participation, so our estimates conflate the effect of involvement with the effect of recognition and should be read as a conservative lower bound. Third, preparation precedes awards, so our post-award coefficients capture the combined effect of the competition experience rather than the effect of the credential alone. Fourth, our administrative data lack measures of motivation, study time, and psychological constructs, so we can only infer mechanisms indirectly. Fifth, three further constraints deserve mention. Our data cover a single cohort in a single discipline, so we cannot separate cohort-specific shocks from stable features of the competition system. The awards in our register are predominantly team-based—the median team size is five and 78 % involve more than one member—and our design attributes the award to every team member, so we cannot distinguish the returns to genuine individual contribution from those to membership of a successful team; the subgroup estimates for national competitions also rest on only 1,781 observations, which is why we do not interpret the negative point estimate for that group substantively. Finally, the register records awards formally reported to the Academic Affairs Office, and any award not reported is invisible to us.

9. Conclusion

This study asked whether competition awards in Chinese higher education certify pre-existing ability or create new capability, using four years of course-level administrative records for 418 undergraduates linked to the timing of their awards. The evidence points to a third answer. Selection into awards is strong and predates treatment—winners were already 0.44 SD ahead of their peers in first-year achievement—so the award itself cannot account for the achievement premium its holders enjoy. What modest post-award movement exists is distributed in a compensatory rather than a cumulative pattern, concentrated among students in the lowest tercile of prior achievement and in specialised coursework. And the sharpest contrast in the data concerns credential quality rather than award frequency: participation prizes that require no competitive success carry no measurable premium, whereas competitively earned awards do. Because the design can reliably detect only effects of roughly 0.07 SD or larger, our inferences rest on the selection gradient and the credential-quality contrast, both of which are precisely estimated, rather than on the magnitude of the average effect.

The theoretical implication is that framing awards as either signals or investments misses what they do in this setting. Awards neither certify an advantage that already exists nor generate large new advantages; they accompany a modest re-engagement with coursework concentrated among students for whom conventional instruction is working least well. We are mindful of the limits of this conclusion: pre-award trends are not perfectly parallel, the average effect is small and fragile to specification, and our data come from a single institution in one disciplinary domain. Future research using multi-institution data, records of participation as well as awards, and measures of the competencies competitions are intended to develop would help establish whether the compensatory pattern we document generalises. Notwithstanding these limitations, this study provides what we believe to be the first course-level event-study evidence on the academic consequences of competition awards in Chinese higher education, and the first to distinguish genuine awards from participation credentials.

These findings carry three practical implications. First, competition programmes should not be justified on the expectation that they raise course grades: the average return we estimate is approximately 0.4 percentage points, small enough that an institution reviewing its competition budget on achievement grounds alone would struggle to defend it. What our results do support is a targeted rather than universal rationale. Because gains concentrate among students in the lowest tercile of prior achievement and are roughly four times larger than for the highest tercile, competition involvement appears to function

as an alternative route into academic engagement for students whom conventional coursework serves least well. Institutions seeking distributional returns should direct outreach, mentoring and preparation support toward these students rather than treating competitions as enrichment for those already succeeding—in our setting, by actively recruiting students who are not in the top tier of their cohort rather than selecting teams on the basis of grades, which is the common practice.

Second, credential quality deserves explicit attention in the rules that govern how awards are used. We find no achievement premium whatsoever for students whose only credential is a participation prize requiring no competitive success, whereas competitively earned awards carry a modest premium. Admissions and scholarship rules that award points on an undifferentiated basis—that is, without distinguishing selective awards from certificates of participation—are therefore weighting credentials that carry no observable academic signal. Distinguishing the two in the points schedule is a low-cost administrative change with a clear evidentiary basis; whether it would also reduce incentives to accumulate credentials of low informational value is a question our data cannot settle but which policy should consider.

Third, the pre-award decline we document deserves attention from those who organise preparation. Students who go on to win score lower in the two to three terms preceding the award, and the most intensive competitors—those reaching national level—show if anything a negative post-award estimate. We do not interpret this as a reason to discourage competition, since the same students recover and the average effect is positive. But it does suggest that preparation schedules competing directly with assessed coursework carry a cost, and that institutions which run intensive training programmes should consider how that load interacts with the teaching calendar.

Statements and Declarations

Ethics approval: The study used anonymised administrative records that cannot be traced to identifiable individuals. In accordance with institutional regulations governing research on such records, the study was exempted from formal ethics review.

Consent to participate: Not applicable. The study used anonymised administrative records that cannot be traced to identifiable individuals.

Data availability: The de-identified data, a codebook describing all variables and the construction of the analytic sample, and the analysis scripts sufficient to reproduce every table and figure reported here are provided as Supplementary Materials accompanying this article and are released under the CC BY 4.0 licence. The crosswalk between anonymised identifiers and original student numbers is retained by the authors and is not part of the package.

Competing interests: The authors declare no competing interests.

Supplementary information: The replication package contains three anonymised data files (data_students.csv, data_scores.csv, data_awards.csv), a codebook describing all variables and sample construction, and the analysis scripts (preprocess.py, build_award.py, analysis.py, analysis_did.py, make_figures.py, make_figures2.py) that reproduce every table and figure reported in the article. Appendix A: Sample Construction Table 5. Sample construction: from raw grade records to the analytic sample.

References

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics, 2*, 193–216.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297–308.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40*, 518–529.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Columbia University Press.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics, 119*, 249–275.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*, 200–230.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47*, 1–32.

Coleman, J. S. (1961). *The adolescent society: The social life of the teenager and its impact on education*. Free Press of Glencoe.

Cooper, H., Valentine, J. C., Nye, B., & Lindsay, J. J. (1999). Relationships between five after-school activities and academic achievement. *Journal of Educational Psychology, 91*, 369–378.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*, 2964–2996.

Eccles, J. S., & Barber, B. L. (1999). Student council, volunteering, basketball, or marching band: What kind of extracurricular involvement matters? *Journal of Adolescent Research, 14*, 10–43.

Fredricks, J. A., & Eccles, J. S. (2006). Extracurricular involvement and adolescent adjustment: Impact of duration, number of activities, and breadth of participation. *Applied Developmental Science, 10*, 132–146.

Fredricks, J. A., & Eccles, J. S. (2008). Participation in extracurricular activities in the middle school years: Are there developmental benefits for African American and European American youth? *Journal of Youth and Adolescence, 37*, 1029–1043.

Fu, X., & Li, Y. (2025). Exploring the behavioural mechanism of college students' participation in academic competition: A structural equation model approach. *European Journal of Education, 60*, e12825.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*, 254–277.

Guest, A., & Schneider, B. (2003). Adolescents' extracurricular participation in context: The mediating effects of schools, communities, and identity. *Sociology of Education, 76*, 89–109.

Guo, F., & Zhang, L. (2023). Participation or award? Research on the educational functions of the college student competition. *China Higher Education Research, 39*, 62–68. (in Chinese)

Hou, H., & Zhao, C. (2026). Computer science competition awards in Chinese higher education: A systems-thinking analysis of expansion, stratification, and portfolio configurations. *Systems, 14*, 871.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*, 241–253.

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities.

Mahoney, J. L., Cairns, B. D., & Farmer, T. W. (2003). Promoting interpersonal competence and educational success through extracurricular activity participation. *Journal of Educational Psychology, 95*, 409–418.

Marsh, H. W., & Kleitman, S. (2002). Extracurricular school activities: The good, the bad, and the nonlinear. *Harvard Educational Review, 72*, 464–515.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics, 37*, 187–204.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). Jossey-Bass.

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies, 90*, 2555–2591.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics, 235*, 2218–2244.

Ruan, Y., Zhang, J., Wang, J., Jian, R., Mei, F., Li, H., Zhang, Y., Hu, Q., Xiao, L., Yang, Y., Li, M., Xiong, J., & Tian, Y. (2025). Academic competition-based learning cultivates scientific literacy to promote professional competitiveness in medical undergraduates. *Frontiers in Public Health, 13*, 1590832.

Schultz, T. W. (1961). Investment in human capital. *American Economic Review, 51*, 1–17.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*, 355–374.
Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*, 175–199.
Yang, L., Liu, Z., & Yin, H. (2026). Examining the relationship between undergraduates' motivation, engagement and learning outcomes in China: A three-wave longitudinal investigation. *Assessment & Evaluation in Higher Education*. Advance online publication. https://doi.org/10.1080/02602938.2026.2656282
Zhou, Q., Hu, S., & Nian, C. (2025). Exploring the impact of public health-related academic competitions on the competencies of university students: Evidence from Anhui Province. *Frontiers in Public Health, 13*, 1600566.

Step	Records	Action
0	47,636	Raw grade records extracted
1	−482	Exclude invalid records (absent, deferred, disqualified, cheating)
2	—	Convert 1,958 letter grades to percentage (95 / 85 / 75 / 65)
3	−28	Drop zero scores (almost certainly non-submission)
4	−694	Restrict to the eight regular terms of the cohort
5	−731	For retaken courses, retain first attempt only
6	−1,338	Drop course-term cells with fewer than ten enrollees
=	44,363	Analytic sample, full eight-term panel (93.1% retention)
7	−4,730	Exclude terms 7–8 (preferred analysis window)
=	39,633	Terms 1–6 descriptive sample (Table 1)
8	−36	Singleton course × term cells dropped in estimation
=	39,597	Preferred estimation sample (terms 1–6; Tables 2–4)