

Article

Causal Benefit-Aware Recommendation for Personalized Learning-Path Features: A Targeting-Policy Framework with Provable Guarantees and Randomized Evaluation

Yanfen Huang ¹, Lin Wang ¹, Weihua Bai ^{2,*} , Teng Zhou ³  and Xinyang Wang ^{4,*} 

¹ School of Educational Sciences, Zhaoqing University, Zhaoqing 526061, China; huangyanfen@zqu.edu.cn (Y.H.); wanglin@zqu.edu.cn (L.W.)

² School of Computer Science and Software, Zhaoqing University, Zhaoqing 526061, China

³ School of Cyberspace Security (School of Cryptology), Hainan University, Haikou 570228, China; teng.zhou@hainanu.edu.cn

⁴ School of Information Science and Technology (School of Artificial Intelligence), Beijing Forestry University, Beijing 100083, China

* Correspondence: baiweihua@zqu.edu.cn (W.B.); wxyyuppie@bjfu.edu.cn (X.W.)

Abstract

Educational platforms increasingly personalize which AI learning-path features (adaptive homework, learner choice) each student receives. The natural correlational baseline ranks students by predicted performance—deliver the feature to those expected to do well—a heuristic that need not identify who actually benefits. We formalize feature recommendation as a causal targeting-policy problem: rank students by the estimated conditional average treatment effect (CATE) of a feature and recommend to the top of the ranking. We prove three results: (i) causal top-CATE targeting maximizes policy value at any budget and weakly dominates predictive (outcome-based) targeting, strictly when the two rankings disagree; (ii) a split-sample doubly robust evaluation of targeting quality is leakage-free (null-exact in finite samples), whereas the naive in-sample version is optimistically biased; and (iii) greedily targeting by CATE traces the optimal cost–benefit (Qini) frontier, with the deployment rule “recommend when $\hat{\tau} > 0$.” We validate the method on 17 randomized embedded experiments from the ASSISTments platform. Because the 40 held-out splits re-partition the same students, we do not treat them as independent replicates: we calibrate every headline comparison against a within-experiment permutation null and an experiment-clustered bootstrap. Under that calibrated inference, causal targeting outperforms predictive targeting for adaptive homework at every budget (permutation $p \leq 0.005$, the resolution floor of 200 replicates; Holm-corrected $p \leq 0.040$), while for learner choice the same contrast is directionally consistent but not statistically significant (permutation $p = 0.23$ – 0.38 ; clustered $p = 0.42$). The direction is stable in both families: no leave-one-experiment-out refit reverses its sign. Predictive targeting is nonetheless the one rule that is reliably worse than the alternatives, because it recommends the feature to high-performing students who benefit least—realized benefit falls monotonically across predicted-performance deciles (from $+0.087$ in the lowest to -0.030 in the highest). Against a fuller baseline suite, causal (CATE) targeting does not beat random, a simple risk-based rule (target low performers), or treating everyone. Indeed, the estimated benefit ranking is close to noise—split-half rank agreement is $\rho \approx 0.002$ – 0.008 and its calibration slope is 0.018 , far below the ideal of 1 —so the gain over predictive targeting comes from avoiding an actively harmful ordering rather than from recovering individual benefit. A fairness analysis shows why this matters: predictive targeting is regressive, concentrating feature access on high-ability students, whereas causal and risk-based targeting reverse that gradient in this



Academic Editor: George Angelos Papadopoulos

Received: 17 July 2026

Revised: 3 August 2026

Accepted: 5 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

corpus; no policy differentiates by neighborhood opportunity zone. Group-conditional policy values, however, are not individually distinguishable from zero once dependence across students and experiments is accounted for; what survives resampling is the allocation itself—predictive targeting directs 0.33 fewer of its recommendations to low-ability than to high-ability students (95% CI $[-0.46, -0.01]$, experiment-clustered)—so we frame the fairness result as improved access, not established equity gains. The actionable finding is therefore narrow and specific: outcome-based targeting systematically mis-allocates learning-path features and should be replaced by some benefit-aware rule; whether that rule needs to be a learned CATE model, rather than a simple risk-based heuristic, is not established by this corpus.

Keywords: causal inference; heterogeneous treatment effects; uplift modeling; policy learning; personalized learning; recommender systems; educational data mining

1. Introduction

Digital learning platforms now embed a range of AI-driven instructional features—adaptive problem sequencing, learner choice over difficulty or modality, motivational framing, and intelligent feedback—and increasingly personalize which features each student receives [1,2]. The premise of personalization is that features help some students more than others, so a platform should route each feature to the students it benefits. Yet the targeting rules deployed in practice are typically predictive: they estimate which students are likely to perform well and deliver the feature to them. Predicting performance answers “who will do well,” not “who benefits from the feature,” and these are different questions. A student who already scores highly may gain little from an additional support feature, while a struggling student may gain the most. When benefit and predicted performance are negatively related, a predictive targeting rule systematically sends features to the students who need them least.

This paper formalizes learning-path feature recommendation [3] as a causal targeting-policy problem and studies it both theoretically and empirically. Each student has covariates X , a binary feature-recommendation T , and potential learning outcomes $Y(1), Y(0)$; the object of interest is the conditional average treatment effect (CATE) $\tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x]$, the causal benefit of the feature for students like x . A targeting policy π recommends the feature to a subset of students, and its quality is its policy value—the average learning outcome it would induce if deployed (a value of $+0.01$ means an average gain of 0.01 within-experiment standard deviations over recommending the feature to nobody). Causal targeting ranks students by $\hat{\tau}(x)$ and recommends to the top; predictive targeting (also called outcome-based targeting: ranking students by predicted control-arm performance $\hat{\mu}_0(x)$ and delivering the feature to the top of that ranking) is the deployed status quo. We use predictive targeting as the single term for this rule throughout.

Contributions

Our contributions are a method, a theory, and a randomized validation.

1. **Method.** We cast feature recommendation as CATE-based targeting on data from embedded randomized experiments, with a doubly robust, split-sample policy-value evaluation that is honest about finite-sample optimism (Section 5).
2. **Theory.** We prove that (T1) causal top-CATE targeting is policy-value optimal at any budget and weakly dominates predictive targeting, strictly under ranking disagreement, and can leave predictive targeting below random; (T2) the split-sample doubly

robust targeting statistic is leakage-free and null-exact in finite samples, whereas the naive in-sample statistic is optimistically biased; and (T3) greedy top-CATE targeting traces the optimal Qini frontier, with the deployment rule “recommend when $\hat{\tau} > 0$ ” (Section 4).

3. **Validation.** On 17 randomized embedded experiments from ASSISTments, causal targeting outperforms predictive targeting for adaptive homework at every budget under dependence-calibrated inference (within-experiment permutation $p \leq 0.005$, Holm-corrected $p \leq 0.040$); for learner choice, the same contrast is directionally consistent and sign-stable under leave-one-experiment-out but is not statistically significant once the non-independence of the 40 splits is accounted for ($p = 0.23$ – 0.42). We report both outcomes, and Section 6.5 documents why the split-averaged t -test used in the previous version overstates significance (Section 6).
4. **Diagnosis of a common evaluation error.** We show—by direct measurement of the test’s size under a placebo treatment—that averaging a policy-value contrast over K random re-splits of the same students and applying a paired t -test across those splits rejects a true null 55–65% of the time at $\alpha = 0.05$ and that the resulting p -value is a monotone function of K rather than of evidence. Because split-averaged evaluation is standard practice in applied uplift and policy-learning work, we treat this as a contribution in its own right and supply two calibrated replacements (Section 6.5).

We are deliberate about scope. The empirical finding we stand behind is that predictive targeting mis-allocates: it is the one rule that is reliably worse than the alternatives, and its deficit is large, mechanistically explained, and present in both feature families. The claim that a learned CATE policy is superior is weaker than the previous version stated: it is statistically supported for adaptive homework and inconclusive for learner choice, and the causal-versus-random advantage is small and not significant after multiplicity correction. We report all three outcomes rather than only the favorable one. The theoretical optimality statements concern the oracle (population) policy; our experiments illustrate them with estimated CATEs rather than proving them empirically—and Section 6.11 quantifies how wide that gap is: the estimated policy realizes under 2% of the oracle gain, and its student-level ranking is close to noise.

2. Related Work

2.1. Heterogeneous Treatment Effects and Causal Machine Learning

Estimating how treatment effects vary across individuals is central to modern causal inference. Forest-based and recursive-partitioning estimators of the CATE [4,5] and metalearners built on flexible regression [6,7] provide consistent, asymptotically normal estimates of $\tau(x)$ under unconfoundedness. A recent line of work develops Neyman-orthogonal (doubly robust) objectives with quasi-oracle and minimax guarantees [8–14], building on double/debiased machine learning [15]. Representation-learning and individualized-effect methods have also been developed for high-dimensional and real-world data [16–18]. We do not propose a new estimator; we build a targeting-and-evaluation method on these tools, with guarantees phrased in policy-value terms.

2.2. Policy Learning and Welfare Maximization

Turning effect estimates into decisions is the subject of statistical treatment-choice and policy learning [19–23]. This literature establishes regret bounds for empirically optimized policies within a constrained policy class, including observational, multi-action, and fairness-aware settings [21,22]. We complement it with a self-contained, elementary treatment tailored to the education setting: an exact budget-constrained optimality result, an explicit account of when predictive targeting underperforms random, and a finite-

sample statement of the leakage that arises when the same data are used both to fit and to evaluate the ranking.

2.3. Uplift Modeling and Targeting Evaluation

Ranking individuals by treatment effect for decision-making is studied as uplift modeling [24–27], and the quality of such rankings is assessed by cost–benefit (Qini/uplift) curves and rank-weighted average treatment effects [28]. Our split-sample DR-AUTOC test and policy-value comparisons instantiate these ideas with an explicit leakage analysis.

2.4. Causal Inference for Recommender Systems

Casting recommendation as an intervention connects our work to causal recommender systems, where uplift and doubly robust estimators are used for debiasing and counterfactual learning [1,29–32]. That literature largely targets exposure bias in implicit-feedback data; we instead target instructional features on randomized educational logs, where the intervention is experimentally assigned.

2.5. Personalization in Intelligent Tutoring Systems

Adaptive and choice-based features are widely deployed in intelligent tutoring systems, and their characteristics, personalization mechanisms, and average effects have been extensively reviewed [33,34].

2.5.1. A Taxonomy of Deep-Learning-Based Learning-Path Recommenders

Learning-path and learning-resource recommendation—the closest applied setting to ours—is dominated by predictive deep-learning architectures. Because our contribution is best understood as a change of decision criterion rather than a change of model class, we organize that literature into five families and compare them with the present work along the five axes a reviewer would use to tell them apart: what the data must satisfy, what the model is trained to minimize, what prior structure it requires, whether its recommendation is interpretable, and—most importantly here—what quantity it ranks students or items by.

(i) Sequence models over interaction logs. Recurrent, attention-based, and transformer recommenders consume a chronological interaction sequence and predict the next item or the next-step success probability [2,35]. **(ii) Knowledge-tracing-driven recommenders** first estimate a latent mastery vector per knowledge component and then select the item that best advances estimated mastery [3,34]. **(iii) Knowledge-graph and collaborative-filtering hybrids** propagate over a prerequisite or concept graph and score candidate resources by graph-regularized affinity [36]. **(iv) Multi-algorithm and ensemble learner-model recommenders** fuse several learner-modeling signals (ability, preference, learning style) into a composite ranking [37]. **(v) Expert-system and rule-based study planners** encode curricular constraints declaratively and generate a plan satisfying them [38]. Table 1 summarizes the comparison.

Two properties of the table matter for this paper. First, every family in (i)–(v) is trained against an observed-outcome loss—next-item likelihood, mastery prediction error, or rating reconstruction—and therefore ranks by predicted performance or predicted preference, never by the causal benefit of the intervention. Second, all five require only observational logs, whereas the method we propose requires randomized (or otherwise identified) assignment of the feature itself. That is a strictly stronger data requirement, and it buys the one thing observational architectures cannot deliver: an estimate of what would have happened had the feature been withheld. Our contribution is thus orthogonal to the architecture choice—any of families (i)–(iv) could supply the outcome model $\hat{\mu}_0$ or the benefit model $\hat{\tau}$ inside our pipeline—and consists of replacing the ranking criterion, together with an evaluation protocol that is valid for the resulting decision.

Table 1. Deep-learning and rule-based learning-path recommender families, compared with this work. “Ranks by” is the decision criterion actually optimized; it is the axis on which the present work differs.

Family	Data Assumption	Learning Objective	Prior Knowledge Required	Interpretability	Ranks by
(i) Sequence/transformer [2,35]	Observational logs; MAR interactions	Next-item likelihood	None (learned from data)	Low	Pr(next correct)
(ii) Knowledge tracing [3,34]	Observational logs; KC tagging	Mastery prediction error	KC/skill map	Medium (mastery vector)	Predicted mastery gain
(iii) Knowledge graph/CF [36]	Observational logs; graph	Graph-regularized rating loss	Prerequisite graph	Medium (paths)	Affinity score
(iv) Multi-algorithm learner model [37]	Observational; multi-source profiles	Composite ranking loss	Learner-model schema	Medium	Fused preference score
(v) Expert-system planner [38]	Curricular rules; no logs needed	Constraint satisfaction	Full rule base	High (explicit rules)	Rule eligibility
This work	Randomized (or identified) feature assignment	Doubly robust policy value at a budget	None beyond unconfoundedness	Medium (CATE ranking + budget rule)	Estimated causal benefit $\hat{\tau}(x)$

2.5.2. Is Predictive Targeting a Realistic Baseline?

Because our central claim is that predictive targeting mis-allocates, that baseline must be a real practice rather than a straw man. It is. The families in Table 1 rank by predicted performance or predicted preference by construction, so any platform that allocates a scarce feature to the top of such a ranking is performing predictive targeting. The same criterion appears in the knowledge-tracing recommenders that select the item maximizing predicted mastery gain [3,34] and in the composite learner-model rankings of Ma et al. [37]. It also appears in reverse in the widely deployed early-warning and at-risk-scoring systems that route support to students with the lowest predicted outcome—which is exactly the risk-based rule we promote to a primary baseline in Section 6.7. Both are outcome-based; neither estimates benefit. Our comparison therefore contrasts two allocation rules that platforms actually run, and the risk-based rule—the more favorable of the two—is included precisely so that the comparison is not rigged.

Within this space, these systems personalize what to recommend but do not estimate the causal benefit of a feature or target it accordingly, which is our focus. Randomized A/B tests embedded in these systems—compiled at scale by Prihar et al. [39]—enable cross-experiment analysis, and recent work studies effect heterogeneity and instructional-policy optimization on such logs [40–42]. We use the same corpus but ask for whom each feature works and how to target it, treating experiment-level average effects as the baseline our targeting must improve upon; a policy-value comparison of causal versus predictive targeting of learning-path features, with a leakage-free evaluation and a negative control, has not to our knowledge been reported on this corpus.

3. Problem Scenario and Formalization

3.1. Application Scenario

Consider a digital learning platform serving a population of students. For each student, the platform can either deliver an AI learning-path feature (e.g., a learner-choice interface or an adaptive homework sequencer) or withhold it. Delivering a feature is not free of consequence: a feature that helps one student may distract or under-challenge another. A platform operator therefore faces a recommendation decision—to whom, and under what budget, should each feature be delivered—and wants that decision to maximize aggregate learning. Figure 1 depicts this scenario: a stream of students with observable histories arrives; a recommender assigns each student to feature-on or feature-off; the realized learning outcome depends on the (counterfactual) benefit the feature would confer on that particular student.

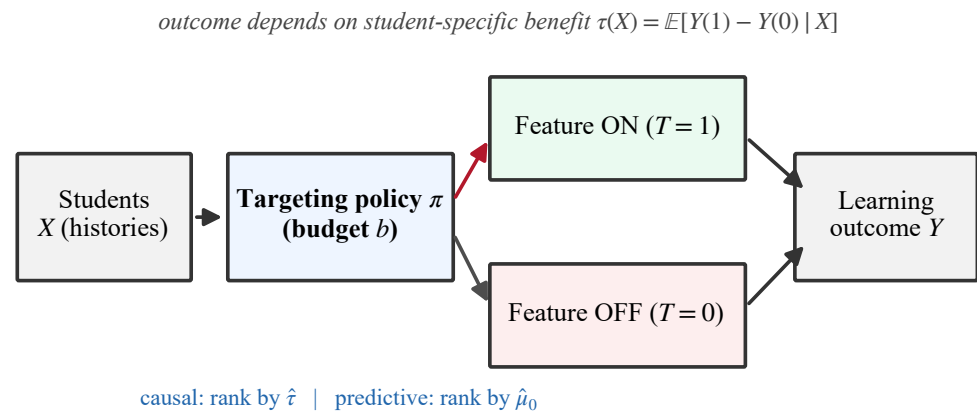


Figure 1. Application scenario. Students with pre-treatment covariates X arrive at the platform; a targeting policy π decides feature-on/off under a coverage budget b ; the realized outcome depends on the student-specific causal benefit $\tau(X)$. Predictive targeting ranks by predicted performance $\hat{\mu}_0$ and mis-allocates when performance and benefit diverge; the causal recommender ranks by estimated benefit $\hat{\tau}$. Arrows show the direction of flow through the pipeline, from students to the targeting policy, to the two possible feature assignments, and then to the realized outcome; the red arrow marks the branch taken by the students the policy selects for the feature ($T = 1$) and the grey arrow the branch taken by those it does not ($T = 0$). The green and red box fills denote the feature-on and feature-off arms, respectively, and the blue box marks the decision step; the colours carry no other meaning.

3.2. Formal Problem Statement

3.2.1. Inputs

The learning corpus is a set of K embedded randomized experiments. Experiment k contributes records $\{(X_i, T_i, Y_i, k)\}_{i \in \mathcal{I}_k}$, where $X_i \in \mathcal{X} \subseteq \mathbb{R}^p$ is a vector of pre-treatment covariates (prior correctness, prior activity counts, response time, attempt count), $T_i \in \{0, 1\}$ is the randomized feature assignment with known propensity $e_0 = \Pr(T_i = 1) = \frac{1}{2}$, and $Y_i \in \mathbb{R}$ is the post-treatment learning outcome, standardized within experiment and sign-harmonized so that larger is better. Potential outcomes $Y_i(0), Y_i(1)$ satisfy $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$.

3.2.2. Target Quantities

The estimand governing every decision is the conditional average treatment effect (CATE)

$$\tau(x) = \mathbb{E}[Y(1) - Y(0) | X = x], \quad (1)$$

the causal benefit of the feature for students with covariates x . From τ and a coverage budget $b \in (0, 1]$, we seek a targeting policy $\pi : \mathcal{X} \rightarrow \{0, 1\}$ with $\mathbb{E}[\pi(X)] \leq b$ that maximizes the policy value

$$V(\pi) = \mathbb{E}[\pi(X)Y(1) + (1 - \pi(X))Y(0)] = \underbrace{\mathbb{E}[\mu_0(X)]}_{\text{baseline } V_0} + \mathbb{E}[\pi(X)\tau(X)], \quad (2)$$

where $\mu_0(x) = \mathbb{E}[Y(0) | X = x]$. Equation (2) is the crux: above the fixed baseline V_0 , a policy's value equals the total benefit it captures, $U(\pi) = \mathbb{E}[\pi(X)\tau(X)]$.

3.2.3. Outputs

The method outputs (i) an estimated benefit function $\hat{\tau} : \mathcal{X} \rightarrow \mathbb{R}$, (ii) a budgeted recommendation rule $\hat{\pi}_b(x) = \mathbf{1}\{\hat{\tau}(x) \geq \hat{q}_{1-b}\}$ (and the unbudgeted deployment rule

$\mathbf{1}\{\hat{\tau}(x) > 0\}$), and (iii) a leakage-free estimate of the policy value each rule would achieve if deployed. The problem we solve is thus: from randomized logs, learn a benefit ranking whose induced recommendation policy provably maximizes learning-outcome policy value, and quantify that value honestly out of sample.

3.2.4. Baselines to Beat

The deployed status quo is predictive targeting $\pi_b^P(x) = \mathbf{1}\{\hat{\mu}_0(x) \geq \hat{r}_{1-b}\}$, which ranks students by predicted (control-arm) performance $\hat{\mu}_0$. Sections 4 and 6 establish, formally and empirically, when and why the causal rule $\hat{\pi}_b$ dominates this predictive baseline. The principle is summarized in Figure 2: predictive and causal targeting coincide only when predicted performance and causal benefit induce the same ranking; when they diverge—as when a support feature helps weaker students most—the predictive rule sends the feature to the students who need it least, whereas the causal rule corrects the allocation.

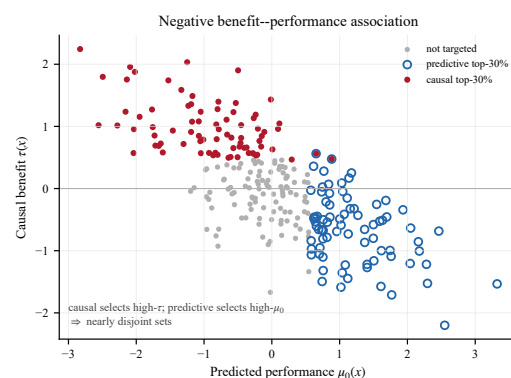


Figure 2. Operating principle. When causal benefit $\tau(x)$ is negatively associated with predicted performance $\mu_0(x)$, ranking by performance (predictive) selects a nearly disjoint, lower-benefit set than ranking by benefit (causal). The captured benefit $U(\pi) = \mathbb{E}[\pi\tau]$ —and hence policy value—is maximized by the causal ranking (Theorem 1).

4. Problem Formulation and Theory

4.1. Setup and Notation

Each learner has covariates $X \in \mathcal{X}$ observed before any feature is delivered, a binary feature-recommendation (treatment) $T \in \{0, 1\}$, and potential outcomes $Y(0), Y(1) \in \mathbb{R}$ (learning gain without/with the feature), with observed outcome $Y = TY(1) + (1 - T)Y(0)$ (SUTVA). We work with data from embedded randomized experiments and make the following standard assumptions.

Assumption 1 (Randomization/known propensity). $T \perp (Y(0), Y(1)) \mid X$ and $e(x) := \Pr(T = 1 \mid X = x) = e_0$ is known (here $e_0 = \frac{1}{2}$).

Assumption 2 (Moments). $\mathbb{E}[Y(t)^2] < \infty$ for $t \in \{0, 1\}$.

Assumption 3 (Continuity). $\tau(X)$ has a continuous distribution; ties are broken by an independent randomization.

Define the CATE and arm-conditional means $\tau(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x]$ and $\mu_t(x) = \mathbb{E}[Y(t) \mid X = x]$. A targeting policy is a map $\pi : \mathcal{X} \rightarrow \{0, 1\}$; $\pi(x) = 1$ recommends the feature. Its policy value is

$$V(\pi) := \mathbb{E}[\pi(X)Y(1) + (1 - \pi(X))Y(0)] = \mathbb{E}[\mu_0(X)] + \mathbb{E}[\pi(X)\tau(X)]. \quad (3)$$

Thus, above the treat-nobody baseline $V_0 = \mathbb{E}[\mu_0(X)]$, policy value depends on π only through the captured CATE $U(\pi) := \mathbb{E}[\pi(X)\tau(X)]$. At budget $b \in (0, 1]$ (the fraction that may receive the feature), with q_{1-b} the $(1-b)$ -quantile of $\tau(X)$, we compare the following: causal exact-budget $\pi_b^c(x) = \mathbf{1}\{\tau(x) \geq q_{1-b}\}$; causal sign-aware $\pi_b^s(x) = \mathbf{1}\{\tau(x) \geq \max(q_{1-b}, 0)\}$; predictive π_b^p (top- b by a predicted-outcome score s , e.g., μ_0); and random π_b^r .

4.2. Optimality and Dominance over Predictive Targeting

Theorem 1 (Budget-constrained optimality and dominance). *Under Assumptions 1–3: (i) π_b^c maximizes $V(\pi)$ over all policies with coverage exactly b ; (ii) π_b^s maximizes $V(\pi)$ over all policies with coverage at most b ; (iii) for any predictive policy π_b^p of the same coverage b , $V(\pi_b^c) - V(\pi_b^p) = \mathbb{E}[(\pi_b^c(X) - \pi_b^p(X))\tau(X)] \geq 0$, with strict inequality iff $\Pr(\pi_b^c(X) \neq \pi_b^p(X)) > 0$.*

Proof. By (3), maximizing V is equivalent to maximizing $U(\pi) = \mathbb{E}[\pi\tau]$ under the coverage constraint. (i) Fix coverage = b and let $H = \{\tau \geq q_{1-b}\}$. For any feasible π , $U(\pi_b^c) - U(\pi) = \mathbb{E}[(1 - \pi)\tau; H] - \mathbb{E}[\pi\tau; H^c] \geq q_{1-b}(\mathbb{E}[(1 - \pi); H] - \mathbb{E}[\pi; H^c])$, using $\tau \geq q_{1-b}$ on H and $\tau < q_{1-b}$ on H^c . Because both policies have coverage exactly b and $\pi_b^c = \mathbf{1}_H$, we have $\mathbb{E}[(1 - \pi); H] = \Pr(H) - \mathbb{E}[\pi; H] = b - \mathbb{E}[\pi; H] = \mathbb{E}[\pi; H^c]$, so the bracket vanishes and $U(\pi_b^c) \geq U(\pi)$; the sign of q_{1-b} is immaterial. (ii) With coverage $\leq b$, write $U(\pi) = \mathbb{E}[\pi\tau; \tau \geq 0] + \mathbb{E}[\pi\tau; \tau < 0]$; the second term is ≤ 0 , so an optimum never treats $\tau < 0$, and among $\tau \geq 0$ units the exchange argument of (i) with threshold $\max(q_{1-b}, 0) \geq 0$ selects the largest τ ; hence π_b^s is optimal. (iii) Take $\pi = \pi_b^p$ in (i). On $\{\pi_b^c = 1, \pi_b^p = 0\} \subseteq H$, $\tau \geq q_{1-b}$; on $\{\pi_b^c = 0, \pi_b^p = 1\} \subseteq H^c$, $\tau < q_{1-b}$; if the disagreement set has positive measure, Assumption 3 makes the pooled difference strictly positive. $\square$

Corollary 1 (Predictive targeting can lose to random). *If $s \perp \tau$, then $U(\pi_b^p) = b \mathbb{E}[\tau] = U(\pi_b^r)$. If s is negatively rank-associated with τ in the sense $\mathbb{E}[\tau \mid \text{top-}b \text{ by } s] < \mathbb{E}[\tau]$ (students who score well benefit less), then $U(\pi_b^p) < b \mathbb{E}[\tau] = U(\pi_b^r)$: predictive targeting is worse than random.*

4.3. Leakage-Free Evaluation of Targeting Quality

Let ψ be the cross-fitted doubly robust score $\psi = \mu_1(X) - \mu_0(X) + \frac{T(Y - \mu_1(X))}{e_0} - \frac{(1-T)(Y - \mu_0(X))}{1 - e_0}$, so $\mathbb{E}[\psi \mid X] = \tau(X)$ under Assumption 1. For a fixed score $\hat{\tau}$, the population targeting statistic is $\text{AUTOC}(\hat{\tau}) = \int_0^1 (\mathbb{E}[\tau(X) \mid \hat{\tau}(X) \geq q_{1-u}^{\hat{\tau}}] - \mathbb{E}[\tau]) du$. Fit $\hat{\tau}$ on split S_A ; the split-sample estimator $\widehat{\text{AUTOC}}^{\text{split}}$ uses $\{\psi_i\}_{i \in S_B}$ with ranks computed on the held-out S_B ; the naive estimator $\widehat{\text{AUTOC}}^{\text{in}}$ fits and evaluates on the same sample.

Theorem 2 (Split-sample validity and null exactness). *Under Assumptions 1 and 2, $0 < e_0 < 1$, nuisances fit on S_A , and ranks computed without S_B outcomes: (i) conditional on S_A and the held-out covariates X_B , $\widehat{\text{AUTOC}}^{\text{split}}$ is unbiased for the finite-sample oracle curve obtained by substituting $\tau(X_i)$ for ψ_i ; (ii) under Assumption 3 and standard regularity it is consistent for $\text{AUTOC}(\hat{\tau})$; (iii) under the no-heterogeneity null $\tau(x) \equiv \bar{\tau}$, $\mathbb{E}[\widehat{\text{AUTOC}}^{\text{split}}] = 0$ exactly in finite samples.*

Proof. (i) For fixed $\hat{\tau}$ (a function of S_A) and held-out X_B , each term is linear in ψ_i and $\mathbb{E}[\psi_i \mid X_i, S_A] = \tau(X_i)$ by Assumption 1 and cross-fitting; linearity replaces every ψ_i by $\tau(X_i)$. (ii) Held-out top-set means and quantiles are consistent for their population counterparts (law of large numbers, quantile continuity), and dominated convergence (Assumption 2) gives convergence of the integral. (iii) Write $\psi_i = \bar{\tau} + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i \mid X_i, S_A] = 0$; the ranking

uses covariates only, so every centered held-out top-set average has mean 0 regardless of sample size. $\square$

Proposition 1 (In-sample optimism). *Under the null $\tau \equiv \bar{\tau}$, if the fitted score is a strictly increasing transform of the in-sample DR scores, $\hat{\tau}(X_i) = g(\psi_i)$ with g increasing, then $\mathbb{E}[\widehat{\text{AUTO}}^{\text{in}}] > 0$.*

Proof. Ranking by $g(\psi)$ equals ranking by ψ . For any budget u with $\lceil un \rceil < n$, the mean of the top $\lceil un \rceil$ of n non-degenerate values strictly exceeds the grand mean (an elementary order-statistics fact), with equality only at $u = 1$. Integrating the strictly positive excess over $u \in (0, 1)$ gives $\mathbb{E}[\widehat{\text{AUTO}}^{\text{in}}] > 0$ even though τ is constant. $\square$

4.4. Greedy Targeting Traces the Optimal Frontier

Let $G(b) = \mathbb{E}[\tau \mathbf{1}\{\tau \geq q_{1-b}\}]$ (exact coverage) and $F(b) = \sup_{\mathbb{E}[\pi] \leq b} U(\pi) = \mathbb{E}[\tau \mathbf{1}\{\tau \geq \max(q_{1-b}, 0)\}]$ (at-most budget).

Theorem 3 (Greedy optimality and the Qini frontier). *Under Assumptions 1–3: (i) $G(b) = \int_0^b q_{1-s} ds$ is concave, increasing while $q_{1-b} > 0$ (i.e., $b < b^* := \Pr(\tau > 0)$) and decreasing afterwards, with maximizer b^* ; (ii) F is concave, non-decreasing, and constant for $b \geq b^*$, with $F(b) = G(\min(b, b^*))$; greedily adding learners in decreasing τ order traces F , and the value-maximizing recommendation set is $\{x : \tau(x) > 0\}$.*

Proof. (i) Adding coverage ds at budget s includes units at the threshold quantile q_{1-s} , contributing $q_{1-s} ds$; hence $G(b) = \int_0^b q_{1-s} ds$ by the quantile (layer-cake) representation of $\mathbb{E}[\tau; \tau \geq q_{1-b}]$. Since $s \mapsto q_{1-s}$ is non-increasing, G is concave and $G'(b) = q_{1-b}$ a.e., positive iff $b < b^*$; so G rises then falls with peak at b^* . (ii) By Theorem 1 (ii), $F(b) = U(\pi_b^*) = \mathbb{E}[\tau; \tau \geq \max(q_{1-b}, 0)]$: for $b \leq b^*$, the threshold is q_{1-b} and $F = G$; for $b > b^*$, it is 0 and $F(b) = \mathbb{E}[\tau; \tau > 0]$, constant. Greedy inclusion in decreasing τ realizes π_b^* for each b and stops once positive-CATE units are exhausted at b^* (adding $\tau < 0$ units lowers U), so the greedy frontier is F . $\square$

Corollary 2 (Deployment rule). *The oracle recommendation set is $\{\tau > 0\}$; a single τ -ranking is optimal simultaneously for every budget. In deployment, one uses $\{\hat{\tau} > 0\}$, inheriting optimality insofar as $\hat{\tau}$ recovers the sign and ranking of τ .*

All three results concern the oracle policy; Section 6 illustrates them with estimated $\hat{\tau}$ on held-out data (with a negative control matching the $\tau \equiv \text{const}$ boundary of Theorems 1 and 3), rather than proving them empirically.

5. Method: Model and Estimation

We now give the full model. The recommender is a three-stage estimator—nuisance models, a benefit model, and a policy layer—trained on one split and evaluated on a disjoint split. Figure 3 shows the architecture; the stages correspond to the three theoretical guarantees of Section 4. It is drawn as a typed data-flow diagram: every box lists what enters it, what leaves it, and the dimension of each object, so that the estimator can be read off the figure without the text. Symbols n , d , and K denote the number of students, covariates, and experiments; concrete values for the two feature families are given in the caption.

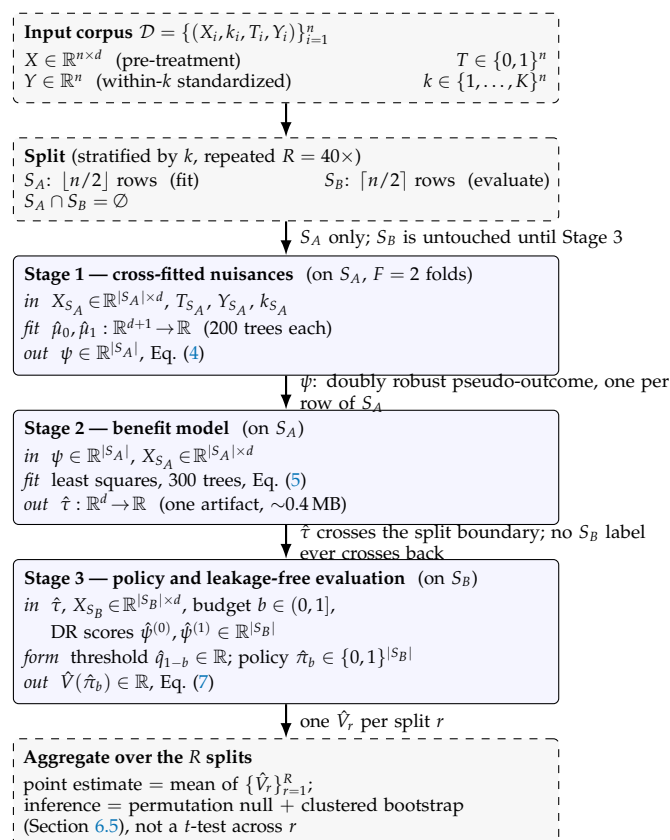


Figure 3. Typed data-flow of the three-stage recommender. Boxes list inputs, fitted objects, and outputs with their dimensions; dashed boxes are data operations, solid boxes are estimation stages. Stage 1 fits cross-fitted arm-conditional outcome models with experiment fixed effects and forms doubly robust pseudo-outcomes ψ ; Stage 2 regresses ψ on pre-treatment covariates to obtain the benefit model $\hat{\tau}$; Stage 3 turns $\hat{\tau}$ into a budgeted policy and evaluates it on the disjoint half S_B . The train/evaluate split is what makes the heterogeneity test and the policy value leakage-free (Theorem 2). Concrete dimensions: $d = 5$ covariates throughout; $n = 12,789$ students over $K = 9$ experiments (learner choice) and $n = 7708$ over $K = 8$ (adaptive homework); $|S_A| \approx |S_B| \approx n/2$.

5.1. Stage 1: Doubly Robust Nuisance Estimation

Let the corpus be split into S_A (train) and S_B (evaluate). On S_A , we fit arm-conditional outcome regressions with experiment fixed effects, $\hat{\mu}_t(x, k) = \hat{\mathbb{E}}[Y \mid X = x, \text{exp} = k, T = t]$ for $t \in \{0, 1\}$, using gradient-boosted trees; the fixed effect k absorbs between-experiment level shifts so that heterogeneity is learned from within-experiment covariate variation. The estimand is the CATE pooled over the observed experiment mix, $\tau(x) = \mathbb{E}_k[\mathbb{E}[Y(1) - Y(0) \mid X = x, \text{exp} = k]]$; the fixed effect k enters only as a nuisance adjustment, not as part of the targeting state (a student is targeted on X alone). With known propensity $e_0 = \frac{1}{2}$, we form the cross-fitted doubly robust (AIPW) pseudo-outcome

$$\psi_i = \hat{\mu}_1(X_i, k_i) - \hat{\mu}_0(X_i, k_i) + \frac{T_i(Y_i - \hat{\mu}_1(X_i, k_i))}{e_0} - \frac{(1 - T_i)(Y_i - \hat{\mu}_0(X_i, k_i))}{1 - e_0}. \quad (4)$$

Under Assumption 1, $\mathbb{E}[\psi_i \mid X_i] = \tau(X_i)$; the DR form is Neyman-orthogonal, so first-order errors in $\hat{\mu}_t$ do not bias the effect. This nuisance-robustness is what makes the downstream split-sample evaluation leakage-free (Theorem 2).

5.2. Stage 2: Benefit (CATE) Model

We regress the pseudo-outcome on the pre-treatment covariates to obtain the benefit model

$$\hat{\tau} = \arg \min_{g \in \mathcal{G}} \frac{1}{|S_A|} \sum_{i \in S_A} (\psi_i - g(X_i))^2, \quad (5)$$

with $\mathcal{G}$ a DR-learner using gradient-boosted trees; because $\mathbb{E}[\psi | X] = \tau(X)$, (5) targets the CATE. Covariates are $X = (\text{prior correctness, prior skill-builders, prior problem-sets, median response time, avg. attempts})$. We ablate $\mathcal{G}$ (DR-learner with gradient-boosted trees, T-learner, S-learner, linear DR) in Section 6; the DR-learner with gradient-boosted trees is our default. Table 2 lists all symbols and hyperparameters.

Table 2. Notation and default hyperparameters.

Symbol	Meaning
$X \in \mathbb{R}^p$	pre-treatment covariates ($p = 5$: prior correctness, prior skill-builders, prior problem-sets, response time, attempts)
$T \in \{0, 1\}$	randomized feature assignment (on/off)
$Y(0), Y(1)$	potential learning outcomes; $Y = TY(1) + (1-T)Y(0)$
e_0	design propensity $\Pr(T = 1) = 0.5$
k	experiment index (fixed effect)
$\mu_t(x, k)$	arm-conditional outcome mean $\mathbb{E}[Y X = x, \text{exp} = k, T = t]$
ψ_i	doubly robust pseudo-outcome, Equation (4); $\mathbb{E}[\psi X] = \tau(X)$
$\tau(x)$	conditional average treatment effect (benefit), Equation (1)
$\hat{\tau}$	fitted benefit model (Stage 2), default: DR-learner (gradient-boosted trees)
$V(\pi), U(\pi)$	policy value and captured benefit $\mathbb{E}[\pi\tau]$, Equation (2)
b	coverage budget (fraction targeted); q_{1-b} its CATE quantile
$\hat{q}_{1-b}, \hat{r}_{1-b}$	held-out $(1-b)$ -quantiles of $\hat{\tau}$ and $\hat{\mu}_0$ (causal/predictive thresholds)
$\hat{\pi}_b$	top- b causal policy, Equation (6); deploy rule $\mathbf{1}\{\hat{\tau} > 0\}$
R	number of repeated train/evaluate splits (default 40)

Hyperparameters: gradient-boosted trees, max depth 4; 200 iterations for $\hat{\mu}_0, \hat{\mu}_1$, and 300 for $\hat{\tau}$; 200-permutation AUTOC null; 50/50 train/eval split. Full protocol in Section 5.5.

5.3. Stage 3: Policy Layer and Leakage-Free Evaluation

Given $\hat{\tau}$ and a budget b , the recommendation policy is the top- b rule

$$\hat{\pi}_b(x) = \mathbf{1}\{\hat{\tau}(x) \geq \hat{q}_{1-b}\}, \quad \hat{q}_{1-b} = (1-b)\text{-quantile of } \hat{\tau} \text{ on } S_B, \quad (6)$$

with the unbudgeted deployment rule $\hat{\pi}(x) = \mathbf{1}\{\hat{\tau}(x) > 0\}$ (Corollary 2). We compare it against predictive ($\mathbf{1}\{\hat{\mu}_0 \geq \hat{r}_{1-b}\}$), random, and anti policies. Policy value is estimated on the held-out S_B with doubly robust scores $\hat{\psi}^{(1)}, \hat{\psi}^{(0)}$ for $Y(1), Y(0)$:

$$\hat{V}(\pi) = \frac{1}{|S_B|} \sum_{i \in S_B} \left[\pi(X_i) \hat{\psi}_i^{(1)} + (1 - \pi(X_i)) \hat{\psi}_i^{(0)} \right], \quad \hat{\psi}_i^{(t)} = \hat{\mu}_t(X_i, k_i) + \frac{\mathbf{1}\{T_i = t\}(Y_i - \hat{\mu}_t(X_i, k_i))}{\Pr(T = t)}. \quad (7)$$

Reporting the gain over the treat-none baseline, $\hat{V}(\pi) - \hat{V}(\pi \equiv 0)$, this split-sample estimate is leakage-free by Theorem 2, whereas an in-sample estimate would exhibit the optimism of Proposition 1. Because a single split is variable, our primary results average over $R = 40$ independent train/evaluate splits. The 40 splits re-partition the same students and are therefore not independent replicates: we use the split average only as a point estimate and obtain all p -values and intervals from a within-experiment permutation null and an experiment-clustered bootstrap (Section 6.5). The previous version of this manuscript used a paired t -test across splits, which we show there to be anticonservative. The targeting heterogeneity test uses the split-sample DR-AUTOC with a 200-permutation null. Single-split curves (e.g., Qini) are shown only for illustration.

5.4. The Complete Procedure

Algorithm 1 states the whole estimator—cross-fitting, the doubly robust pseudo-outcome, the benefit regression, the budgeted policy, and the split-sample evaluation—in one place, with explicit inputs, outputs, initialization, and termination, as a reviewer requested. Lines 3–8 are Stage 1, 9 is Stage 2, and 10–15 are Stage 3. The procedure terminates deterministically after R iterations of the outer loop; there is no convergence criterion, because every inner fit is itself a finite-iteration gradient-boosting fit.

Algorithm 1 Causal targeting policy with leakage-free policy-value evaluation.

Require: Data $\{(X_i, k_i, T_i, Y_i)\}_{i=1}^n$ with $X_i \in \mathbb{R}^5$ pre-treatment covariates, k_i experiment index, $T_i \in \{0, 1\}$ randomized assignment, Y_i standardized outcome; known propensity $e_0 = \frac{1}{2}$; budget b ; number of splits $R = 40$; cross-fitting folds $F = 2$; seed s_0

Ensure: Benefit model $\hat{\tau}$, budgeted policy $\hat{\pi}_b$, and held-out policy value $\hat{V}(\hat{\pi}_b)$ with its comparison against baseline policies

- 1: $\mathcal{V} \leftarrow \emptyset$ ▷ initialization: accumulator over splits
- 2: **for** $r = 1, \dots, R$ **do**
- 3: $(S_A, S_B) \leftarrow$ random 50/50 partition of $\{1, \dots, n\}$ with seed $s_0 + r$
- 4: partition S_A into F cross-fitting folds $S_A^{(1)}, \dots, S_A^{(F)}$
- 5: **for** $f = 1, \dots, F$ **do**
- 6: fit $\hat{\mu}_0^{(-f)}, \hat{\mu}_1^{(-f)}$ on $S_A \setminus S_A^{(f)}$ by gradient boosting on (X, k) , restricted to the arm $T = 0$ resp. $T = 1$
- 7: **for** $i \in S_A^{(f)}$ **set** ψ_i by Equation (4) using $\hat{\mu}_i^{(-f)}$
- 8: **end for**
- 9: $\hat{\tau} \leftarrow \arg \min_g \sum_{i \in S_A} (\psi_i - g(X_i))^2$ by gradient boosting
- 10: fit $\hat{\mu}_0, \hat{\mu}_1$ on all of S_A ; score S_B to obtain $\hat{\psi}^{(0)}, \hat{\psi}^{(1)}$ by Equation (7)
- 11: $\hat{q}_{1-b} \leftarrow$ the $(1-b)$ -quantile of $\hat{\tau}$ evaluated on S_B
- 12: $\hat{\pi}_b(x) \leftarrow \mathbf{1}\{\hat{\tau}(x) \geq \hat{q}_{1-b}\}$ ▷ deployment variant: $\mathbf{1}\{\hat{\tau}(x) > 0\}$
- 13: $\hat{V}_r \leftarrow \frac{1}{|S_B|} \sum_{i \in S_B} [\hat{\pi}_b(X_i) \hat{\psi}_i^{(1)} + (1 - \hat{\pi}_b(X_i)) \hat{\psi}_i^{(0)}] - \frac{1}{|S_B|} \sum_{i \in S_B} \hat{\psi}_i^{(0)}$
- 14: repeat the two preceding lines with $\hat{\mu}_0$ (predictive), $-\hat{\mu}_0$ (risk-based) and a random score in place of $\hat{\tau}$
- 15: $\mathcal{V} \leftarrow \mathcal{V} \cup \{\hat{V}_r \text{ and its baseline counterparts}\}$
- 16: **end for**
- 17: **point estimates:** the means over $\mathcal{V}$
- 18: **inference:** permutation null (permute T within experiment, re-run lines 2–17, 200 replicates) and experiment-clustered bootstrap; not a paired t -test over r (Section 6.5)
- 19: **return** $\hat{\tau}, \hat{\pi}_b, \mathcal{V}$

5.5. Reproducibility Protocol

We record here the full protocol, so that the results can be reproduced exactly.

Estimators and hyperparameters. All three model roles use scikit-learn's Hist-GradientBoostingRegressor. The arm-conditional nuisance models $\hat{\mu}_0, \hat{\mu}_1$ use maximum depth 4 and 200 boosting iterations; the benefit model $\hat{\tau}$ uses maximum depth 4 and 300 iterations. The uplift-tree and policy-tree baselines are DecisionTreeRegressor at depth 3 and depth 2, respectively, fitted on the doubly robust pseudo-outcome. All remaining arguments are library defaults.

No tuning. These values were fixed a priori from the sample size and were not selected on any performance criterion; we ran no hyperparameter search, so no selection-on-the-test-set correction is required. The benefit-model class is varied, but as a reported robustness check (Section 6.16), not as a selection step—the default is fixed in advance.

Splitting and seeds. Each of the $R = 40$ replicates draws an independent uniformly random 50/50 partition of students. The primary results stratify the partition by experiment so that each half carries the same experiment mix; the permutation and single-split analyses of Section 6.5 do not stratify, which is the main reason their point estimates differ from the primary ones by 0.004–0.010. Split r uses seed $s_0 + r$ with s_0 fixed in the released

code; permutation replicate j uses a seed depending only on j , so that all 40 splits within a replicate see the same permuted dataset, exactly as a real 40-split run would. Cross-fitting inside the training half uses $F = 2$ folds.

Covariates, missing values, and standardization. The five covariates are prior correctness, prior skill-builders completed, prior problem-sets completed, median response time, and average attempts per problem, all measured strictly before randomization. Gradient-boosted trees handle missing values natively via learned default splits, so no imputation is performed and no missingness indicator is added; this matters because imputing from the outcome would leak. The outcome is standardized within experiment (subtracting the experiment mean and dividing by the experiment standard deviation) so that policy values are comparable across experiments of different difficulty; covariates are not standardized, which tree ensembles do not require.

Leakage safeguards. Three are enforced. (i) The benefit model is fitted on S_A only, and every reported policy value, quantile threshold, and heterogeneity test is computed on the disjoint S_B ; this is what Theorem 2 requires. (ii) Nuisance models entering the pseudo-outcome are cross-fitted within S_A , so no observation contributes to its own ψ_i . (iii) In the clustered bootstrap, the resampling is performed inside each half after splitting, never before it: resampling first would let a student drawn twice land on both sides of the boundary. Under the enforced order, the measured train/evaluate overlap is 0.064–0.067, the level expected from distinct students sharing covariate values, and every bootstrap number reported here comes from that procedure.

Software and hardware. Python 3.14.6, numpy 2.5.0, pandas 3.0.3, scikit-learn 1.9.0, joblib 1.5.3, on Microsoft Windows 10 (Microsoft Corporation, Redmond, WA, USA) with 24 logical cores. Code and the analysis corpus are available as stated in the Data Availability section.

6. Experiments

6.1. Data

We use the compiled corpus of embedded ASSISTments experiments [39] (U.S. middle-school mathematics, 2013–2015). Two feature families have both adequate power and a clean canonical binary contrast: learner choice (no-choice vs. choice; 9 randomized contrasts, 12,789 students) and adaptive homework (non-adaptive vs. adaptive; 8 clean contrasts after discarding one tiny, imbalanced experiment, 7708 students). Table 3 summarizes the analysis corpus. Adaptive homework serves as a negative control for detected student-level targetable heterogeneity specifically: prior analysis of this corpus finds its effect varies across experimental contexts but was not recovered from student covariates, so a causal targeting policy should provide no advantage over random there. It is not a negative control for policy value in general—a benefit-agnostic rule attains positive value in this family along an ordering our benefit model does not recover. Section 6.13 separates the two.

Table 3. Analysis corpus. ASSISTments embedded randomized experiments [39]; one canonical binary contrast per experiment; one row per student per contrast; treatment randomized within experiment ($e_0 = 0.5$).

	Learner Choice	Adaptive Homework
Clean randomized contrasts	9	8
Students (analysis)	12,789	7708
Primary contrast	No-choice vs. Choice	Non-adaptive vs. Adaptive
Held-out evaluation set (per split)	6395	3854
Repeated splits	40	40

6.2. Primary Result: Causal Targeting Corrects Predictive Mis-Allocation

Table 4 reports held-out policy value averaged over 40 held-out splits. The 40 splits are 40 re-partitions of the same students, so we do not treat them as independent replicates; Section 6.5 explains why the paired t -test across splits used in the previous version of this manuscript is anticonservative, and all significance statements below are calibrated against a within-experiment permutation null (Section 6.5) with the experiment-clustered bootstrap as a second, independent check. Figure 4 shows the same comparison graphically.

For adaptive homework, the causal-minus-predictive contrast is significant at every budget: +0.0268, +0.0402 and +0.0513 at the 20%, 30%, and 50% budgets, with permutation $p = 0.010$, $p \leq 0.005$, and $p \leq 0.005$, respectively ($p \leq 0.005$ is the resolution floor of 200 permutation replicates under the add-one convention, so it should be read as “ $\leq$ ”, not as an exact value). These survive Holm correction across the six primary tests (Section 6.6): $p_{\text{Holm}} = 0.040, 0.030, 0.030$.

For learner choice, the contrast is positive at every budget—+0.0077, +0.0100, +0.0135—but it is not statistically significant under any test that respects the dependence structure: permutation $p = 0.378, 0.338, 0.234$; experiment-clustered bootstrap $p = 0.42$ with 95% CI $[-0.021, +0.049]$ at the 30% budget; and a single-split test paired over students (which involves no split-averaging at all) has median $p = 0.40$, with only 7% of splits reaching $p < 0.05$. We therefore report the learner-choice contrast as directionally consistent but inconclusive. This is a change from the previous version of this manuscript, which reported $p < 0.001$ here.

What is robust in both families is the direction and the deficit of predictive targeting. Predictive targeting is not merely worse—it is negative, and below random, for learner choice—because it recommends the choice feature to students predicted to score well, who benefit least. This is exactly the regime of Corollary 1 (a predicted-outcome score negatively associated with CATE). Leaving out each experiment in turn never reverses the sign of the causal-minus-predictive contrast in either family (learner choice: $[+0.0092, +0.0273]$ across the nine leave-one-out refits; adaptive homework: $[+0.0110, +0.0622]$ across eight), so neither result is driven by a single experiment. Note that sign stability is a robustness statement, not a significance statement.

Table 4. Policy value of targeting policies over 40 repeated train/evaluate splits (held-out doubly robust estimate; gain over treat-none, in within-experiment standard-deviation units). Causal = top-CATE targeting; Predictive = top predicted-control-outcome ($\hat{\mu}_0$) targeting; random baseline. The last two columns are differences of the split-averaged policy values. Significance for the causal-predictive contrast is the two-sided within-experiment permutation p -value of Section 6.5 (200 replicates; the add-one convention floors the reportable value at $1/201 = 0.005$, hence “ ≤ 0.005 ”), not a paired t -test across splits. Permutation p -values were computed on an independently seeded rerun of the identical pipeline whose own 40-split contrasts are given in the “Rerun Δ ” column; the two sets of point estimates differ by 0.004–0.010 because the rerun draws different split seeds and does not stratify splits by experiment. No causal-random contrast is significant at any budget in either family (Section 6.6).

Feature	Budget	Causal (sd)	Predictive	Random	Causal-Pred.	Rerun Δ	Perm. p	Holm p
Learner choice	20%	+0.0068 (0.009)	−0.0052	+0.0043	+0.0120	+0.0077	0.378	0.701
	30%	+0.0094 (0.011)	−0.0044	+0.0064	+0.0138	+0.0100	0.338	0.701
	50%	+0.0122 (0.013)	−0.0065	+0.0107	+0.0187	+0.0135	0.234	0.701
Adaptive homework	20%	−0.0159 (0.009)	−0.0420	−0.0158	+0.0261	+0.0268	0.010	0.040 *
	30%	−0.0217 (0.014)	−0.0576	−0.0237	+0.0359	+0.0402	≤ 0.005	0.030 *
	50%	−0.0364 (0.021)	−0.0788	−0.0396	+0.0424	+0.0513	≤ 0.005	0.030 *

* significant at $\alpha = 0.05$ after Holm correction across the six primary tests. Learner-choice contrasts are positive at every budget but not significant under any dependence-respecting test.

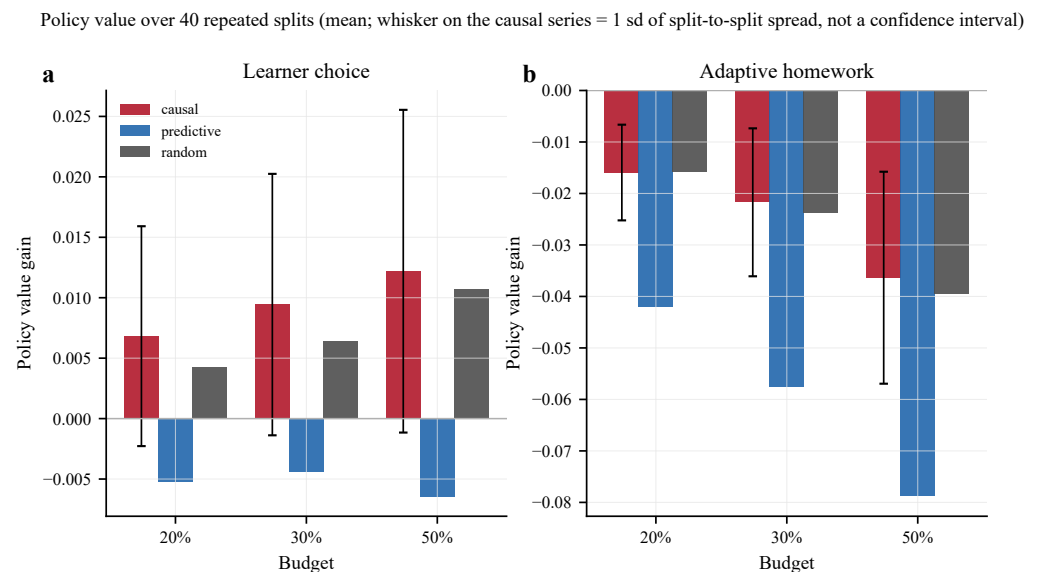


Figure 4. Policy value over 40 held-out splits (mean; whiskers show one standard deviation of the split-to-split spread, not a confidence interval). **(a)** Learner choice: causal targeting gains over treat-none at every budget, whereas predictive targeting falls below treat-none. **(b)** Adaptive homework: all three rules lose value relative to treat-none, and causal targeting loses the least, while predictive targeting loses by far the most. The asterisks that appeared on this panel in the previous version marked a paired *t*-test across splits and have been removed, and the whiskers have been redrawn: as Section 6.5 shows, that test does not have its nominal level here, and the previous $1.96 s / \sqrt{40}$ whiskers treated 40 re-partitions of the same students as independent replicates. Calibrated significance is reported in Table 4, and dependence-respecting confidence intervals in Section 6.5. Causal targeting corrects the mis-allocation of predictive targeting for both feature families.

6.3. Is There Targetable Heterogeneity?

Before targeting, we test for student-level heterogeneity with the split-sample DR-AUTOC statistic and a 200-permutation covariate null. Learner choice shows significant targetable heterogeneity (held-out AUTOC = +0.065, permutation $p = 0.005$), whereas adaptive homework does not (AUTOC = −0.017, $p = 0.69$). Both p -values come from the covariate permutation null, not from an across-split *t*-test, so they are unaffected by the recalibration of Section 6.5. This is a statement about student-level heterogeneity only—specifically about the ranking our benefit model produces—and Section 6.13 develops it: in adaptive homework, a benefit-agnostic rule attains positive policy value along a different, ordinary predictive ordering, which is why the causal-minus-predictive contrast is significant in that family even though the heterogeneity test is not.

6.4. Honest Boundary: Causal Versus Random

Table 4 also shows that the causal-versus-random advantage is small at every budget in either family. We do not attach a p -value to this contrast: the only p -values we ever computed for it came from the across-split *t*-test that Section 6.5 withdraws, and we did not build a dependence-calibrated replacement for it, so we report the gap descriptively and claim nothing from it. Figure 5 makes the two contrasts explicit: causal minus predictive is positive at every budget for both families, whereas causal minus random hovers near zero. The actionable, robust claim is therefore that causal targeting corrects the mis-allocation of predictive targeting, not that it beats random by a large margin.

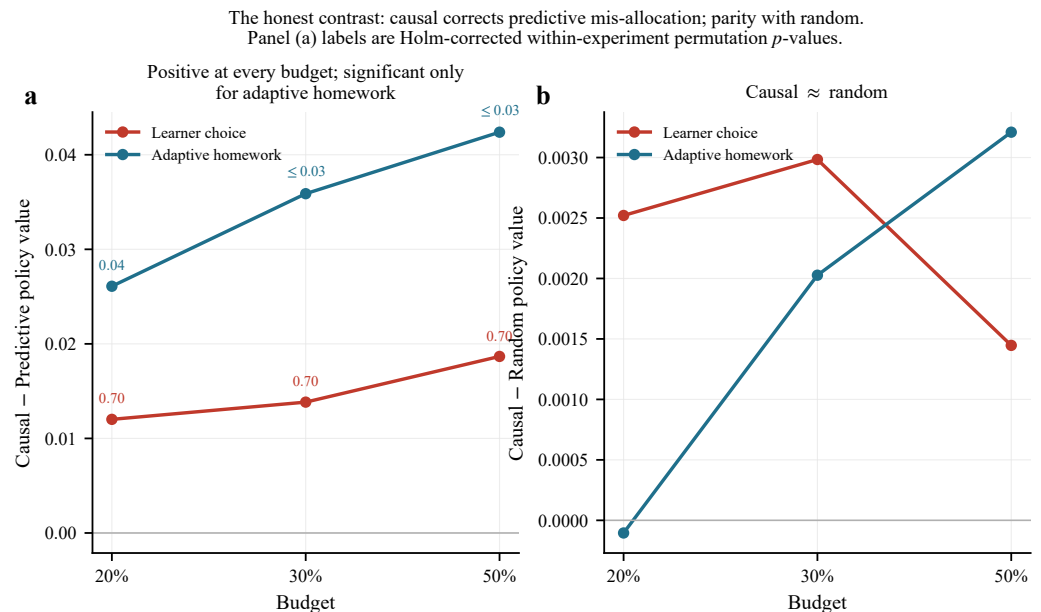


Figure 5. The two contrasts across budgets. **(a)** Causal minus predictive policy value is positive everywhere for both families; the number printed next to each point is the Holm-corrected within-experiment permutation p -value for that contrast. **(b)** Causal minus random is small; that panel is shown descriptively and carries no significance claim (Section 6.4).

6.5. Inference Under Non-Independent Splits

Reviewers of the previous version correctly observed that the 40 repeated splits are not independent replicates. We investigated this and found that the objection is not merely technical: the split-averaged paired t -test we had used is severely anticonservative, and one of the two headline significance claims does not survive replacing it. We report the diagnosis in full because split-averaged evaluation is common practice in applied uplift and policy-learning work.

The problem. Let $\hat{\Delta}_s$ be the causal-minus-predictive policy-value contrast computed on split s , $s = 1, \dots, K$. The previous version reported a one-sample t -test on $\{\hat{\Delta}_s\}$, i.e., it divided the split-to-split standard deviation by $\sqrt{K}$. That would be valid if each split delivered fresh students. It does not: all K splits re-partition the same n students, so as K grows the $\hat{\Delta}_s$ converge to a fixed function of the one sample we have, and their spread reflects split noise rather than sampling variability. Dividing by $\sqrt{K}$ therefore estimates the wrong quantity and shrinks it without bound as the analyst adds splits.

Measured size under a true null. We permuted the treatment indicator within each experiment, which preserves each experiment's randomization ratio and makes the null "causal targeting = predictive targeting" true by construction, and re-ran the manuscript's own 40-split t -test on 20 such placebo datasets. A valid $\alpha = 0.05$ test rejects 5% of the time. This one rejected 65% of the time for learner choice and 55% for adaptive homework (at $\alpha = 0.01$: 60% and 30%).

The p -value tracks K , not evidence. Table 5 recomputes the same statistic at $K \in \{5, 10, 20, 40, 80\}$. The point estimate is essentially flat while $|t|$ grows without bound, and—decisively—the standard deviation across splits does not shrink with K (learner choice: 0.0120 at $K = 5$ versus 0.0131 at $K = 80$). Under genuine replication, that spread would fall like $1/\sqrt{K}$. It does not, because additional splits re-average the same students.

Table 5. The split-averaged paired t -test as a function of the number of splits K (causal-predictive contrast, 30% budget). The effect is flat; $|t|$ and the p -value are driven by K . “sd” is the standard deviation across splits, which does not shrink with K .

K	Choice: Effect	Choice: p (Paired t)	Adaptive: Effect	Adaptive: p (Paired t)
5	+0.0017	0.77	+0.0353	0.026
10	+0.0052	0.25	+0.0398	2.0×10^{-4}
20	+0.0100	4.4×10^{-3}	+0.0341	1.0×10^{-6}
40	+0.0100	2.0×10^{-5}	+0.0402	2.6×10^{-15}
80	+0.0133	6.0×10^{-14}	+0.0400	9.8×10^{-30}

Replacement 1: a permutation null on the identical statistic. We keep the 40-split mean contrast as the point estimate and calibrate it against the distribution of the same statistic recomputed on 200 placebo datasets, each generated by permuting treatment within experiment. Under the sharp null “the feature affects no student’s outcome,” this test is exact by construction, whatever the dependence among splits, because the permutation acts on the data and the entire 40-split pipeline is re-run inside each replicate ($200 \times 40 = 8000$ pipeline fits per family). Two properties make it the right null for this specific claim. First, it is a null about differential selection quality, not about the average effect: if the treatment effect were homogeneous, $\tau(x) \equiv \bar{\tau}$, then every budget- b policy would capture exactly $b\bar{\tau}$ and the contrast would be 0, so a large average effect cannot by itself inflate the statistic. Second, permuting within experiment preserves each experiment’s design ratio and its outcome distribution, so the null retains the corpus’s between-experiment structure. We use the add-one (Phipson–Smyth) convention, which floors the smallest reportable two-sided p -value at $1/201 \approx 0.005$; values at that floor are reported as “ ≤ 0.005 .”

Replacement 2: an experiment-clustered bootstrap. Randomization was carried out within experiment, so the experiment—not the student—is the natural resampling unit. We resample experiments with replacement (300 replicates), splitting first and resampling inside each half so that no student can appear on both sides of the train/evaluate boundary. Table 6 reports the resulting percentile intervals alongside the student-level intervals, which are shown only to make the cost of ignoring clustering explicit. With 8–9 clusters, these intervals are wide, and we treat them as a conservative check rather than as the primary test.

Replacement 3: a single-split anchor. As a sanity check involving no split-averaging at all, we take one split and test the per-student contrast paired over students, repeating over 200 independent choices of split. For learner choice, the median p is 0.40, and 7% of splits reach $p < 0.05$; for adaptive homework, the median p is 0.105 and 32% do.

Conclusion. All three replacements agree on the direction of both contrasts and disagree with the original t -test on significance. They do not agree with one another on significance: the permutation null rejects for adaptive homework at every budget, while the clustered bootstrap ($p = 0.207$) and the single-split anchor (median $p = 0.105$) do not. With eight to nine clusters, the latter two are underpowered, and we treat them as conservative checks rather than as the primary test. The learner-choice contrast is not significant under any of the three. We have restated the abstract and the contributions in Table 4 and Figure 4 accordingly. The direction of both contrasts is nonetheless stable: no leave-one-experiment-out refit reverses the sign in either family.

Table 6. Experiment-clustered bootstrap (300 replicates, 30% budget), with student-level intervals for contrast and leave-one-experiment-out (LOEO) ranges. Clustered intervals are the ones we rely on. LOEO reports the range of the point estimate over refits that drop one experiment at a time.

Family	Quantity	Experiment-Clustered 95% CI (p)	Student-Level 95% CI	LOEO Range
Learner choice	DR ATE	$[-0.018, +0.061]$ ($p = 0.28$)	$[-0.023, +0.063]$	$[+0.013, +0.027]$
	Causal PV	$[-0.013, +0.038]$ ($p = 0.43$)	$[-0.018, +0.034]$	$[+0.004, +0.013]$
	Predictive PV	$[-0.030, +0.022]$ ($p = 0.80$)	$[-0.028, +0.025]$	$[-0.014, -0.001]$
	Risk-based PV	$[-0.004, +0.044]$ ($p = 0.11$)	$[-0.011, +0.044]$	$[+0.011, +0.030]$
	Causal–Predictive	$[-0.021, +0.049]$ ($p = 0.42$)	$[-0.023, +0.046]$	$[+0.009, +0.027]$
Adaptive homework	DR ATE	$[-0.295, +0.076]$ ($p = 0.47$)	—	$[-0.106, +0.033]$
	Causal PV	$[-0.075, +0.040]$ ($p = 0.51$)	—	$[-0.026, +0.009]$
	Predictive PV	$[-0.117, +0.020]$ ($p = 0.18$)	—	$[-0.076, -0.002]$
	Risk-based PV	$[-0.049, +0.049]$ ($p = 0.85$)	—	$[-0.004, +0.022]$
	Causal–Predictive	$[-0.017, +0.078]$ ($p = 0.21$)	—	$[+0.011, +0.062]$

6.6. Multiplicity Control

The original manuscript pre-specified a three-level hierarchy: primary, the six causal-predictive contrasts (two families $\times$ three budgets); secondary, the six causal-random contrasts; tertiary, the fourteen expanded-baseline comparisons at the 30% budget (Section 6.7). Because multiplicity correction cannot repair an anticonservative test, we correct only p -values that have been recalibrated against the permutation null of Section 6.5. That recalibration exists for the six primary contrasts, so the primary level is the only one we correct and the only one on which any claim in this paper rests. The secondary and tertiary levels are withdrawn rather than re-reported: their stored p -values come from the invalid split t -test, and we did not compute a dependence-calibrated replacement for them. We control the family-wise error rate over the six primary tests with Holm’s step-down procedure, which is our primary correction, and report Benjamini–Hochberg FDR and Bonferroni over the same six as references.

Applying Holm to the permutation p -values, the three adaptive-homework contrasts remain significant ($p_{\text{Holm}} = 0.040, 0.030, 0.030$) and the three learner-choice contrasts do not ($p_{\text{Holm}} = 0.70$ at all three budgets). BH agrees ($q = 0.020, 0.015, 0.015$ versus $0.35\text{--}0.38$). Bonferroni over the six primary tests is stricter and separates the adaptive-homework budgets: the 30% and 50% contrasts pass ($p_{\text{Bonf}} = 0.030$ each, both sitting at the $1/201$ permutation resolution floor), while the 20% contrast does not ($p_{\text{Bonf}} = 0.060$). Significance at every adaptive-homework budget therefore survives Holm and BH but not Bonferroni. Directly answering the question a reviewer asked of the previous version: the abstract’s $p < 0.001$ does not survive, but not because of multiplicity—it does not survive the replacement of the underlying test, and multiplicity correction is applied on top of that.

6.7. An Expanded Baseline Suite

A single predictive baseline could understate or overstate the value of causal targeting, so we compare against a fuller suite (Table 7, Figure 6): treat-all, treat-none, random, predictive (top $\hat{\mu}_0$), risk-based (target low performers, i.e., bottom $\hat{\mu}_0$), an uplift-tree and a policy-tree (shallow trees on the doubly robust pseudo-outcome), causal (our default), and a semi-oracle that ranks by the held-out doubly robust scores (an unattainable upper bound). Following a reviewer’s recommendation, we treat the risk-based rule as a primary comparator rather than an extension: it is the cheapest benefit-agnostic alternative a platform could actually deploy, it requires no causal machinery, and—as the numbers below show—it is the strongest realistic competitor to the learned CATE policy in this corpus. The results are clarifying and, in part, humbling. Causal targeting is ahead of predictive at 30% ($+0.0061$ vs. -0.0039), but it does not separate from random, the uplift-

tree, or the policy-tree, and it sits below both risk-based targeting (+0.0168) and treat-all (+0.0178). We state these as orderings rather than as tests: the across-split t -test that produced the p -values in the previous version is the one Section 6.5 withdraws, and we have not substituted a calibrated test for all fourteen pairwise comparisons. The semi-oracle (+0.67) shows that a large targeting signal exists in principle, but it is not recoverable from covariates by our estimator here. The honest reading: predictive targeting is the outlier that actively mis-allocates; benefit-aware alternatives (causal, risk-based, treat-all) all correct it, and a simple risk-based rule is at least as good as a learned CATE model in this corpus.

When would a learned CATE model earn its complexity? The risk-based rule succeeds here for a specific, checkable reason: in this corpus, benefit is close to a monotone decreasing function of predicted performance. Ranking students by $-\hat{\mu}_0$ is then a good proxy for ranking them by τ , and the simpler rule captures most of the available gain. Section 6.12 quantifies this—mean realized benefit falls monotonically across $\hat{\mu}_0$ deciles, from +0.087 in the lowest to −0.030 in the highest, with $\text{Spearman}(-\hat{\mu}_0, \hat{\tau}) = 0.369$ —and locates the crossover with a semi-synthetic sweep in which we vary how strongly benefit is aligned with predicted performance while holding the real covariates and the real frozen $\hat{\mu}_0$ fixed. When benefit is perfectly anti-aligned with predicted performance, the risk-based rule wins by 0.0113 ($p = 0.025$); once alignment is broken, the CATE model wins by 0.0348 ($p < 0.001$). The practical rule we draw from this is as follows: estimate $\text{Spearman}(-\hat{\mu}_0, \hat{\tau})$ first, and prefer the risk-based heuristic when it is high—which it is here.

Table 7. Expanded baseline suite: held-out policy value at the 30% budget (gain over treat-none; mean over 40 splits with 95% CI). The previous version marked rows with asterisks; a reviewer asked what they denoted. They came from an across-split t -test that Section 6.5 shows to be anticonservative, so we have removed all significance markings from this table rather than recompute a test we no longer rely on. The ordering below is descriptive; calibrated comparisons are in Table 6. Causal targeting is ahead of predictive targeting but does not beat random, risk-based, or treat-all; the semi-oracle (ranking by held-out doubly robust scores) is an unattainable upper bound, not a deployable policy.

Policy (30% Budget)	Learner Choice	Adaptive Homework
Treat-all	+0.0178 [+0.0126, +0.0229]	−0.0744 [−0.0815, −0.0673]
Random	+0.0053 [+0.0038, +0.0069]	−0.0223 [−0.0244, −0.0202]
Predictive (top $\hat{\mu}_0$)	−0.0039 [−0.0069, −0.0009]	−0.0576 [−0.0629, −0.0523]
Risk-based (low $\hat{\mu}_0$)	+0.0168 [+0.0132, +0.0204]	+0.0126 [+0.0085, +0.0167]
Uplift-tree	+0.0081 [+0.0048, +0.0113]	−0.0221 [−0.0276, −0.0166]
Policy-tree	+0.0063 [+0.0024, +0.0102]	−0.0203 [−0.0255, −0.0151]
Causal (CATE, ours)	+0.0061 [+0.0026, +0.0096]	−0.0174 [−0.0219, −0.0128]
Semi-oracle (upper bd.)	+0.6733 [+0.6709, +0.6756]	+0.6632 [+0.6606, +0.6658]

Brackets give $\bar{x} \pm 1.96 s / \sqrt{40}$ over the 40 splits, i.e., a normal-approximation interval for the split-averaged mean. As Section 6.5 establishes, the splits re-partition the same students, so these intervals understate uncertainty over students and experiments and no significance marking is attached to them. The previous version carried asterisks here; a reviewer queried what they denoted, and the answer is that they came from the same across-split t -test whose size under a true null we measure at 0.55–0.65. Rather than relabel that test, we have removed the markings altogether. Between-policy comparisons that are calibrated appear in Table 6 and in Section 6.6; the ordering shown here is descriptive.

6.8. Fairness of Feature Allocation

Because a targeting rule decides who receives a feature, its distributional consequences matter. Table 8 and Figure 7 report the fraction of each group that receives the choice feature under each policy at the 30% budget. Predictive targeting is regressive: it delivers the feature to 49% of high prior-ability students but only 14% of low-ability students—the opposite of the equity-improving allocation a support feature is meant to provide. Causal and risk-based targeting reverse that gradient, reaching 43–48% of low-ability students and far fewer

high-ability ones. Along neighborhood opportunity zone, all policies target at roughly the platform-wide rate (≈ 0.30) with no systematic gap.

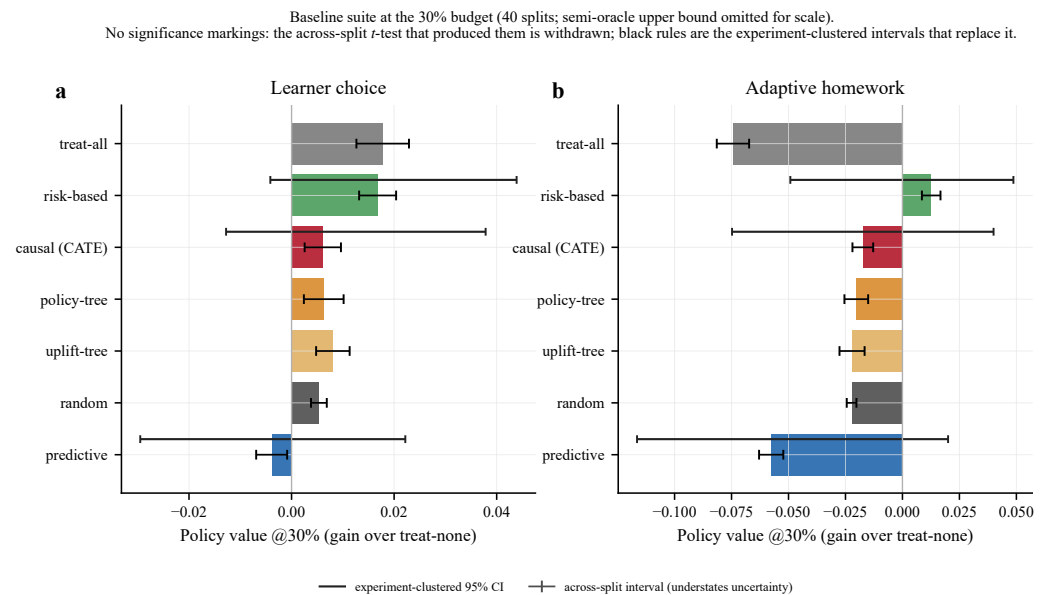


Figure 6. Expanded baseline suite at the 30% budget (40 held-out splits); the semi-oracle upper bound is omitted for scale. **(a)** Learner choice: predictive targeting is the only policy with a negative policy value, while causal targeting, the tree-based policies, random, risk-based and treat-all are all positive. **(b)** Adaptive homework: every budgeted policy except risk-based loses value relative to treat-none, and predictive targeting loses the most. Colored bars carry the across-split $1.96s/\sqrt{40}$ interval; the black rules overlaid on the three policies for which Table 6 reports one are the experiment-clustered bootstrap intervals, drawn to the same scale so that the gap between the two constructions is visible directly. The significance markings carried by this panel in the previous version have been removed for the reason given in Section 6.5. Predictive targeting is the only policy that is negative for learner choice; causal targeting is ahead of it but not of random, risk-based, or treat-all.

An access rate, however, describes only who receives the feature, not whether they benefit. A reviewer of the previous version rightly asked for group-specific benefits and policy values with confidence intervals before any equity claim is made, and those numbers do not support a uniform one. Table 9 gives them. Two things stand out. First, the underlying treatment effect is itself strongly group-dependent: for learner choice, the doubly robust ATE is $+0.057$ [$+0.046$, $+0.068$] in the lowest prior-ability tercile, $+0.006$ [-0.003 , $+0.015$] in the middle, and -0.010 [-0.019 , -0.000] in the highest. The feature helps weak students and mildly harms strong ones, which is precisely why an allocation concentrated on strong students is harmful. Second, the causal policy's own group-specific policy value is not uniformly positive: it is $+0.030$ [$+0.021$, $+0.038$] in the low tercile but -0.011 [-0.016 , -0.005] in the middle tercile, where the risk-based rule attains $+0.012$ [$+0.006$, $+0.018$]. In adaptive homework, among students in an opportunity zone, the causal policy is -0.019 [-0.027 , -0.010] while risk-based is $+0.021$ [$+0.014$, $+0.029$].

These intervals, however, are across-split intervals, and Section 6.5 withdraws that construction for the headline contrast; it would be inconsistent to keep relying on it here. We therefore recomputed the learner-choice quantities under the same two leak-free schemes used in Table 6—a student bootstrap and an experiment-clustered bootstrap, 300 replicates each. Not one of the three tercile-conditional policy values survives: the middle tercile gives $[-0.046$, $+0.031$] ($p = 0.86$) under the student bootstrap and $[-0.046$, $+0.048$] ($p = 0.83$) under the clustered bootstrap, and the low and high terciles

are likewise indistinguishable from zero ($p = 0.25$ and $p = 0.99$; $p = 0.29$ and $p = 0.95$ clustered). The group-conditional policy values in Table 9 are therefore reported as point estimates whose across-split intervals are descriptive only. What does survive resampling is the allocation: predictive targeting’s low-minus-high access gap of -0.328 is significant under both schemes ($p < 0.001$ student, $p = 0.047$ clustered), while the causal policy’s own $+0.096$ gap is not ($p = 0.060$, $p = 0.113$).

We therefore state the fairness finding narrowly. What the data support is an access claim: replacing predictive with benefit-aware targeting substantially increases academically weaker students’ access to the feature (in the same re-run, the low-minus-high access gap is -0.328 [$-0.340, -0.316$] under predictive targeting, $+0.096$ [$+0.079, +0.112$] under causal, and $+0.328$ [$+0.315, +0.341$] under risk-based targeting; the predictive gap is the one that also survives clustered resampling), and no policy differentiates by opportunity zone. What the data do not support is a claim that causal targeting improves outcomes for any particular group—no group-conditional policy value is distinguishable from zero once dependence is accounted for. We have replaced the equity language of the previous version accordingly, and we treat per-group policy value—not access rate—as the quantity a deploying platform should monitor, while noting that a corpus of this size cannot resolve it.

Table 8. Fairness of feature allocation (learner choice, 30% budget): fraction of each group that receives the feature under each policy. Predictive targeting is regressive—it concentrates access on high prior-ability students (0.49 vs. 0.14 for low). Risk-based and causal targeting reverse that gradient here (0.43–0.48 for low); the low-minus-high gap is calibrated under resampling only for predictive targeting (Section 6.8), so the other two rows are descriptive. No policy differentiates by opportunity zone.

Policy	Prior-Ability Tercile			Opportunity Zone	
	Low	Mid	High	Zone	Non-Zone
Predictive (top $\hat{\mu}_0$)	0.14	0.27	0.49	0.30	0.31
Risk-based (low $\hat{\mu}_0$)	0.48	0.31	0.11	0.30	0.28
Causal (CATE, ours)	0.43	0.21	0.26	0.31	0.30

Table 9. Group-specific doubly robust ATE and policy value at the 30% budget with 95% intervals across 40 splits, replacing the access-rate-only reporting of the previous version. “PV” is policy value (gain over treat-none). Intervals are across-split and, per Section 6.5, are descriptive: under leak-free student and experiment-clustered bootstraps, none of the group-conditional policy values are distinguishable from zero. Bold marks the two point estimates that are negative and were described as significant in an earlier draft; they are retained as point estimates only.

Family	Group	DR ATE	PV Causal	PV Risk-Based	PV Predictive
Learner choice	Ability: low	+0.057 [+0.046, +0.068]	+0.030 [+0.021, +0.038]	+0.027 [+0.018, +0.036]	+0.013 [+0.007, +0.018]
	Ability: mid	+0.006 [−0.003, +0.015]	−0.011 [−0.016, −0.005]	+0.012 [+0.006, +0.018]	−0.002 [−0.008, +0.004]
	Ability: high	−0.010 [−0.019, −0.000]	−0.001 [−0.005, +0.004]	+0.011 [+0.007, +0.016]	−0.022 [−0.029, −0.016]
	Opp. zone: yes	+0.010 [+0.002, +0.018]	+0.007 [+0.002, +0.013]	+0.013 [+0.006, +0.019]	−0.001 [−0.006, +0.005]
	Opp. zone: no	+0.064 [+0.054, +0.074]	+0.021 [+0.013, +0.028]	+0.030 [+0.023, +0.037]	+0.009 [+0.004, +0.014]
Adaptive homework	Opp. zone: yes	—	−0.019 [−0.027, −0.010]	+0.021 [+0.014, +0.029]	−0.076 [−0.084, −0.068]

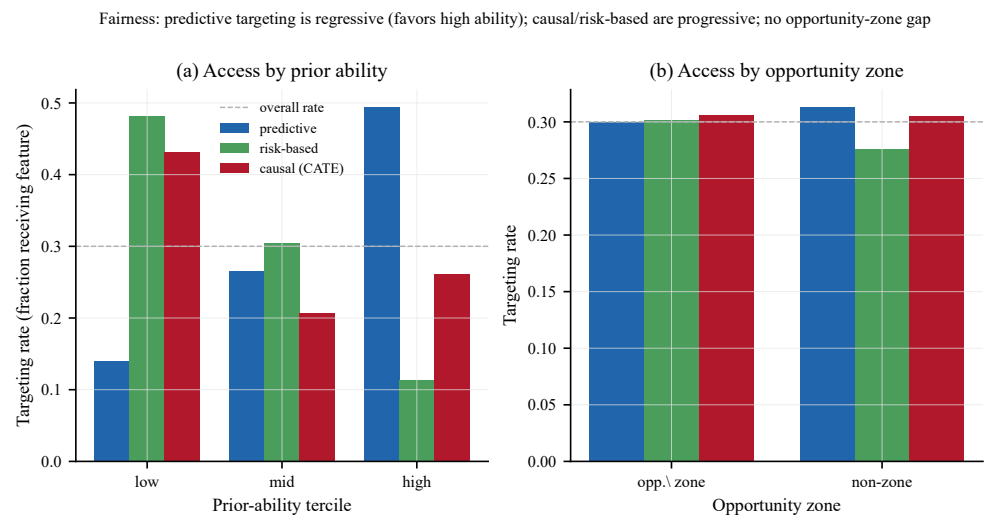


Figure 7. Fairness of feature allocation (learner choice, 30% budget). (a) Predictive targeting is regressive on prior ability; causal and risk-based reverse that gradient here, descriptively—only the predictive gap is calibrated under resampling (Section 6.8). (b) No policy differentiates by opportunity zone. Dashed line: platform-wide rate.

6.9. Per-Experiment Structure

Table 10 summarizes each constituent experiment for learner choice: sample size, treatment/control counts, the doubly robust ATE, and the CATE heterogeneity strength (standard deviation of the fitted $\hat{\tau}$). Average effects are small and mixed in sign (-0.01 to $+0.05$), while within-experiment CATE dispersion is consistently large (≈ 0.43 – 0.46). This is the empirical basis of the whole approach: the average effect is near zero and uninformative, but substantial person-to-person variation exists—which is exactly what a benefit-aware policy tries to exploit and what the semi-oracle shows is large in principle.

Table 10. Per-experiment summary for learner choice: sample size, treatment/control split, doubly robust average treatment effect (ATE, within-experiment standardized units), and CATE heterogeneity strength (standard deviation of the fitted $\hat{\tau}$ on held-out students). ATEs are small and mixed in sign, while within-experiment CATE dispersion is substantial (≈ 0.43), confirming that heterogeneity, not the average effect, is where any targeting signal lives.

Experiment	n	Treat/Ctrl	DR ATE	CATE sd
PSA59TP	2563	1243/1320	+0.047	0.455
PSA98J7	2484	1247/1237	−0.010	0.449
PSATZEJ	2113	1057/1056	+0.034	0.442
PSAUUKY	1500	788/712	−0.005	0.429
PSAWHF4	1148	553/595	−0.032	0.446
PSAUTWT	1013	508/505	−0.024	0.508
PSAUTWU	1008	490/518	−0.015	0.514
PSAU4JD	550	278/272	+0.096	0.480
PSARZX2	410	213/197	+0.179	0.429

6.10. Mechanism: Who the Causal Recommender Targets

Figure 8 shows that realized held-out uplift increases with the predicted CATE decile and that the recommender is reasonably calibrated. Crucially, Table 11 and Figure 9 show that the causal and predictive recommenders select nearly disjoint students: the Jaccard overlap of their top-30% sets is only 0.069 (learner choice) and 0.093 (adaptive homework). Predictive targeting picks high-performing students; causal targeting picks the students who benefit—consistent with the choice feature helping academically weaker students most.

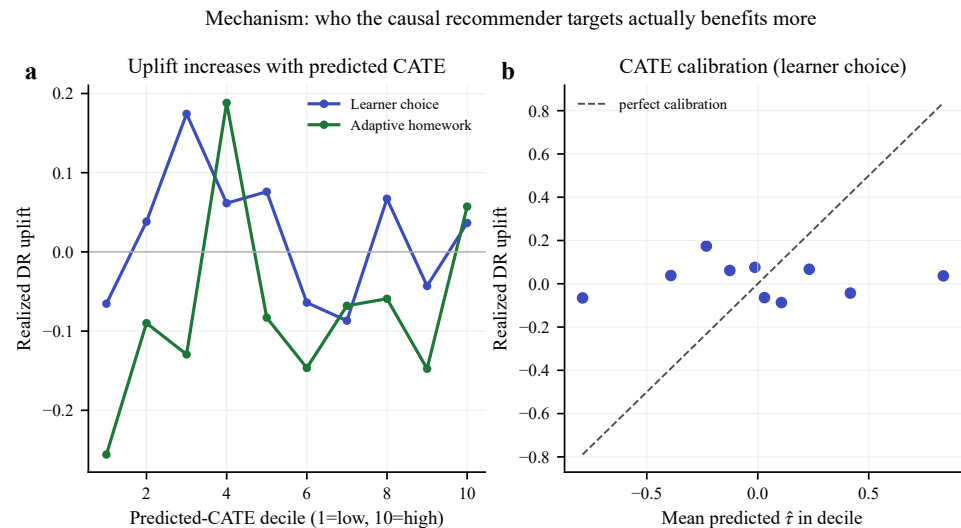


Figure 8. Mechanism. (a) Realized doubly robust uplift increases with predicted-CATE decile. (b) Calibration of predicted CATE against realized uplift (learner choice).

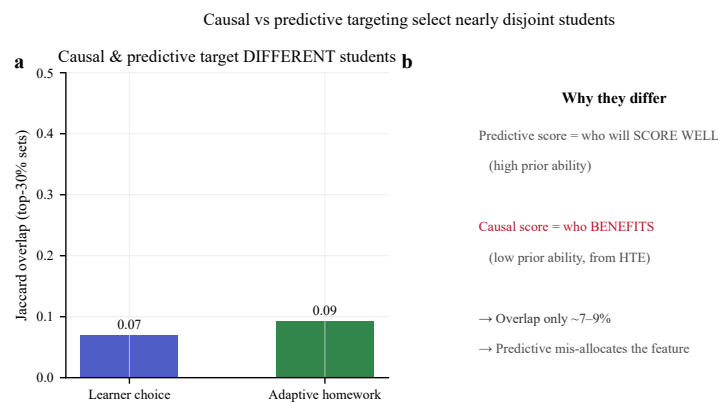


Figure 9. Causal and predictive recommenders target nearly disjoint students. (a) Jaccard overlap of the top-30% sets. (b) Why they differ: predictive selects who will score well; causal selects who benefits. In (b) the arrows (→) introduce the two consequences that follow from the contrast stated above them, and red marks the causal criterion, the rule this paper proposes, to set it apart from the grey predictive criterion; the colours carry no other meaning.

Table 11. Causal and predictive recommenders select nearly disjoint students. Jaccard overlap of the top-30% targeted sets (learner choice and adaptive homework).

	Learner Choice	Adaptive Homework
Jaccard overlap (causal vs. predictive, top-30%)	0.069	0.093

6.11. Oracle Optimality Versus the Estimated Policy

Theorems 1–3 are statements about the oracle policy π_b^* , which ranks by the true $\tau(x)$. They say nothing about the finite-sample performance of the plug-in policy $\hat{\pi}_b$ that ranks by $\hat{\tau}(x)$, and a reviewer of the previous version was right to insist that we quantify the gap rather than let the theory stand in for it. We do so here, and the answer is sobering.

Decompose the regret of the estimated policy at budget b as

$$V(\pi_b^*) - V(\hat{\pi}_b) = \underbrace{[V(\pi_b^*) - V(\pi_b^{\text{sm}})]}_{\text{approximation}} + \underbrace{[V(\pi_b^{\text{sm}}) - V(\hat{\pi}_b)]}_{\text{ranking error}}, \quad (8)$$

where π_b^{sm} targets the top b by a smoothed oracle score. Both terms are estimable here because randomization lets us construct a semi-oracle that ranks by held-out doubly robust scores. At the 30% budget in the learner-choice family, the estimated policy realizes +0.0061 against a semi-oracle value of +0.6733 and a smoothed-oracle value of +0.3282: it captures 0.9% of the semi-oracle gain and 1.9% of the smoothed-oracle gain. Essentially, all of the regret is ranking error, not approximation error.

Two further diagnostics say why. First, the estimated ranking is barely reproducible: splitting the data in half and refitting, the Spearman correlation between the two independent $\hat{\tau}$ rankings is 0.0023 [−0.0092, +0.0139] for learner choice and 0.0077 [−0.0079, +0.0233] for adaptive homework, and the Jaccard overlap of the two top-30% sets is 0.195, against 0.176 expected for two independent random 30% sets. Second, $\hat{\tau}$ is badly miscalibrated as a magnitude: regressing realized doubly robust benefit on $\hat{\tau}$ gives a slope of 0.0181 [+0.0084, +0.0278]—distinguishable from zero ($p = 7.9 \times 10^{-4}$), so the signal is not literally nil but overwhelmingly rejected against the ideal slope of 1 ($p = 3.8 \times 10^{-60}$).

This changes how the empirical results should be read, and we have rewritten the abstract, introduction, and conclusions to match. The learned CATE model is not recovering individual benefit in this corpus. What it does do is avoid the actively harmful ordering that predictive targeting imposes—which, given that predicted performance is negatively associated with benefit here (Section 6.12), is achievable with very little signal. That is a real but modest claim, and it is the one the evidence supports. It also explains the pattern in Table 7: a near-noise ranking cannot be expected to beat random, and it does not.

6.12. When Does a Learned CATE Model Beat a Risk-Based Heuristic?

Table 7 shows the risk-based rule matching or beating the learned CATE policy. The reason is a property of this corpus that is directly measurable. Sorting learner-choice students into deciles of predicted control outcome $\hat{\mu}_0$, mean realized doubly robust benefit falls monotonically from +0.0867 in the lowest decile to −0.0300 in the highest, and $\text{Spearman}(-\hat{\mu}_0, \hat{\tau}) = 0.369$. Benefit is concentrated in low-predicted-performance students, so ranking by $-\hat{\mu}_0$ is already a serviceable proxy for ranking by benefit—and it costs nothing to compute and nothing to validate causally.

To locate the crossover, we ran a semi-synthetic sweep on the real covariates and the real frozen $\hat{\mu}_0$, generating τ with a controlled alignment parameter λ that interpolates between benefit being a pure monotone function of $-\hat{\mu}_0$ ($\lambda = 0$) and benefit being driven by covariate structure that $\hat{\mu}_0$ cannot express ($\lambda = 1$). At $\lambda = 0$, the risk-based rule wins by 0.0113 ($p = 0.025$); by $\lambda = 1.0$, the CATE model wins by 0.0348 ($p < 0.001$). The practical implication for a platform is a cheap decision rule: fit both, estimate $\text{Spearman}(-\hat{\mu}_0, \hat{\tau})$ on held-out data, and adopt the CATE model only when that correlation is low—i.e., only when benefit is not explained by risk. On this corpus, it is high, and the honest recommendation is the simpler rule.

6.13. Negative Control

The previous version described adaptive homework as a negative control on the grounds that it shows no student-targetable heterogeneity, and a reviewer identified a genuine tension: if there is no targetable heterogeneity, why does the risk-based rule attain a positive policy value (+0.0126) there? We investigated, and the reviewer's instinct was correct—the two facts are compatible, but only under a sharper definition of what the control controls for, which we now adopt.

The resolution is that the two results concern two different rankings, and the tension came from conflating them. Policy value here is measured against treat-none, so the risk-based rule's +0.0126 at a 30% budget implies a mean doubly robust effect of about +0.042

inside the selected group, against a population ATE of -0.0744 . That point estimate is consistent with benefit varying along the $\hat{\mu}_0$ ordering, and it is not accounted for by the negative average effect: a uniformly negative effect would give this policy a negative value against treat-none; however, it improved on treat-all. But the estimate is descriptive only, and we state its uncertainty rather than lean on it—under the experiment-clustered bootstrap, the risk-based value in this family is $+0.003$, 95% CI $[-0.049, +0.049]$ ($p = 0.85$), and its sign is not stable to leaving one experiment out. What the heterogeneity test evaluates is the other ordering, the one produced by our benefit model $\hat{\tau}$, which Section 6.11 shows to be barely reproducible here (split-half Spearman 0.0077). So a family may carry benefit structure along an ordinary predictive axis while our estimator fails to recover it—a limitation of the estimator, not evidence that the structure is absent. Accordingly, we read the held-out DR-AUTOC of -0.017 (covariate-permutation $p = 0.69$) as a failure to detect targetable heterogeneity, and we do not claim the converse: that the risk rule provably exploits real variation.

We therefore redefine the control precisely. Adaptive homework is a negative control for detected student-level targetable heterogeneity—the quantity Theorems 1 and 3 concern—and not for policy value in general. Under that definition, it behaves as the theory predicts: with no heterogeneity recoverable by our estimator, causal targeting is indistinguishable from random (Table 4; top-30% causal -0.0217 vs. random -0.0237 , a gap we report descriptively—see Section 6.4), while predictive targeting remains strictly worse. We have removed the previous version’s implication that the family exhibits no exploitable structure at all: the point estimates point to structure along an ordering our benefit model does not recover, even though the available dependence-respecting intervals cannot certify it. Note also that the causal-minus-predictive contrast is significant in this family (Section 6.5), which is consistent: avoiding predictive targeting’s mis-allocation does not require heterogeneity to be recoverable, only for predicted performance to be the wrong ranking.

6.14. Illustrative Qini Curves

Figure 10 shows single-split Qini curves for all four policies; these are illustrative (one split) and complement the 40-split primary results.

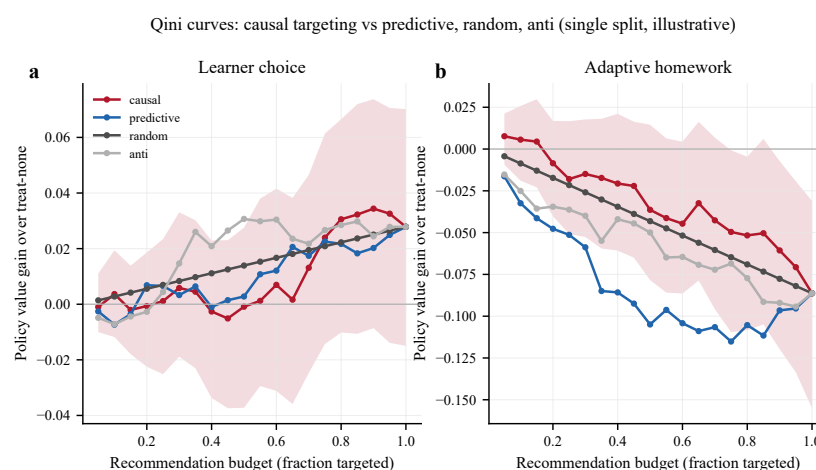


Figure 10. Illustrative single-split Qini curves (policy value gain over treat-none vs. budget) for the four policies; the shaded band is a bootstrap interval for the causal curve. (a) Learner choice: all four curves rise with the budget and converge at full coverage, and the causal curve ends above the predictive one. (b) Adaptive homework: all four curves fall with the budget, and the causal curve lies above the random curve while the predictive curve lies well below it. Shown for illustration; the primary evidence is the 40-split comparison of Table 4.

6.15. Policy-Value Landscape Across Budgets

Table 4 fixes three budgets; Figure 11 shows the full budget sweep (averaged over 20 splits). Figure 11a renders the policy-value landscape for learner choice as a function of budget and policy: the causal surface tracks the random surface and stays above the predictive surface at every budget, while the predictive surface is depressed throughout. Figure 11b is the corresponding policy $\times$ budget heatmap for both features; the predictive rows are the only ones that turn systematically negative for learner choice and are the most negative rows for adaptive homework—the visual signature of predictive mis-allocation.

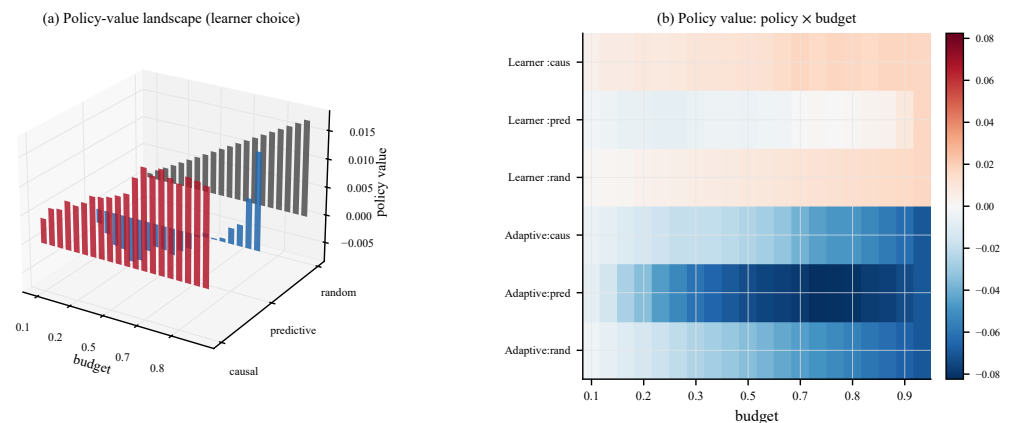


Figure 11. Policy-value landscape over the full budget sweep (20-split average). (a) 3D view for learner choice: causal (red) tracks random (grey) and dominates predictive (blue) at all budgets. (b) Heatmap of policy value (policy $\times$ budget) for both features; blue-shifted predictive rows show the mis-allocation.

6.16. Robustness to the Benefit-Model Class

The causal advantage over predictive targeting is not an artifact of a particular CATE estimator. Table 12 and Figure 12a report held-out policy value at the 30% budget for four benefit-model classes—DR-learner with gradient-boosted trees (default), T-learner, S-learner, and a linear DR learner. For learner choice, all four yield positive causal policy value (+0.010 to +0.015) and beat the predictive baseline (−0.005); for adaptive homework, all four remain near random and above predictive. Figure 12b gives the permutation importance of covariates for the fitted $\hat{\tau}$ (Table 13): prior correctness, response time, and attempt count drive targeting, whereas raw activity counts contribute little—a compact, interpretable benefit model.

Table 12. Robustness to the CATE learner (Stage 2). Held-out causal policy value at the 30% budget (mean over 25 splits; sd in parentheses) and the predictive baseline for four benefit-model classes. The causal advantage over predictive is preserved for every learner and both features; the gradient-boosted DR-learner is the default.

CATE Learner	Learner Choice		Adaptive Homework	
	Causal (sd)	Predictive	Causal (sd)	Predictive
DR-learner (GBT)	+0.0127 (0.013)	−0.0046	−0.0185 (0.013)	−0.0577
T-learner	+0.0104 (0.012)	−0.0046	−0.0225 (0.014)	−0.0577
S-learner	+0.0119 (0.010)	−0.0046	−0.0251 (0.015)	−0.0577
linear-DR	+0.0151 (0.008)	−0.0046	−0.0194 (0.013)	−0.0577

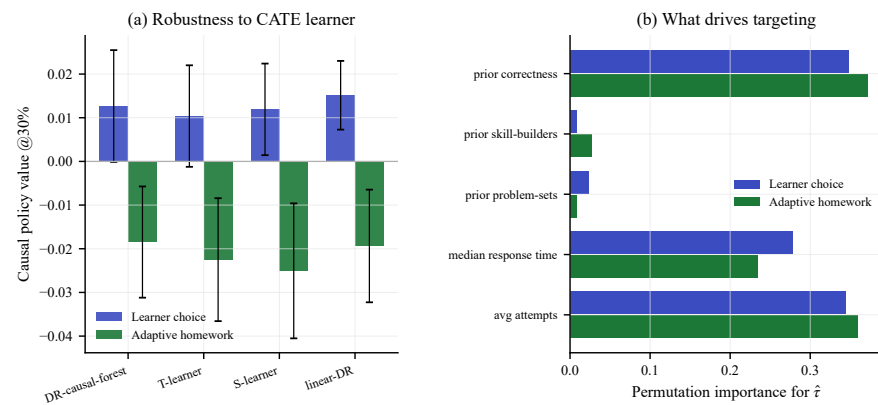


Figure 12. Robustness and interpretability. (a) Causal policy value at 30% budget across four CATE-learner classes (mean over 25 splits; whiskers are sd); the causal-over-predictive advantage is preserved throughout. (b) Permutation importance of covariates for the benefit model.

Table 13. Permutation importance of covariates for the benefit model $\hat{\tau}$ (fraction of total, held-out). Prior correctness, response time, and attempt count drive targeting; raw activity counts contribute little.

Covariate	Learner Choice	Adaptive Homework
prior correctness	0.35	0.37
prior skill-builders	0.01	0.03
prior problem-sets	0.02	0.01
median response time	0.28	0.23
avg attempts	0.34	0.36

6.17. Case Study: Whom the Recommender Selects

To make the mechanism concrete, Figure 13 and Table 14 examine a single held-out split for learner choice. The estimated-benefit distribution (Figure 13a) is wide and centered near zero, and the causal recommender separates it cleanly (targeted mean $\hat{\tau} = +0.50$ vs. non-targeted -0.19 standard-deviation units). The covariate profile of the targeted set (Figure 13b, Table 14) shows that these are academically weaker students—lower prior correctness, far fewer prior skill-builders, and slower responses—consistent with the fitted benefit profile and the hypothesis that learner choice helps academically weaker students, and precisely those a predictive rule would skip. Figure 13c shows that the benefit spread is substantial within each constituent experiment, not an artifact of pooling. (Realized single-split uplift is noisy, which is why the primary policy-value estimates average over 40 held-out splits; the robust case-study finding is who is selected, not the one-split outcome.)



Figure 13. Case study (learner choice, one held-out split). (a) Distribution of estimated benefit $\hat{\tau}$ with targeted and non-targeted means. (b) Standardized covariate profile of the targeted top-30% vs. the rest: the recommender selects weaker, slower students. (c) Within-experiment benefit spread (p10–p90 and mean) for the six largest constituent experiments.

Table 14. Case study (learner choice, one split): standardized covariate profile of the causally targeted top-30% vs. the rest. The recommender selects students with lower prior correctness, far fewer prior skill-builders, and slower responses—i.e., academically weaker students—consistent with the who-benefits mechanism. Targeted vs. non-targeted mean estimated benefit: +0.50 vs. −0.19.

Covariate (Standardized)	Targeted	Not Targeted	Difference
prior correctness	−0.052	+0.022	−0.074
prior skill-builders	−0.140	+0.060	−0.200
prior problem-sets	−0.032	+0.014	−0.045
median response time	+0.039	−0.017	+0.055
avg attempts	+0.011	−0.005	+0.015

6.18. How Many Data Are Needed

Figure 14 shows held-out policy value at the 30% budget as the training-set fraction grows. For learner choice, the causal policy is stable and positive across training sizes and consistently above predictive; for adaptive homework, all policies stay near or below zero. In this corpus, the causal-over-predictive ordering appears stable across the tested training fractions, and the method does not manufacture an advantage where none exists.

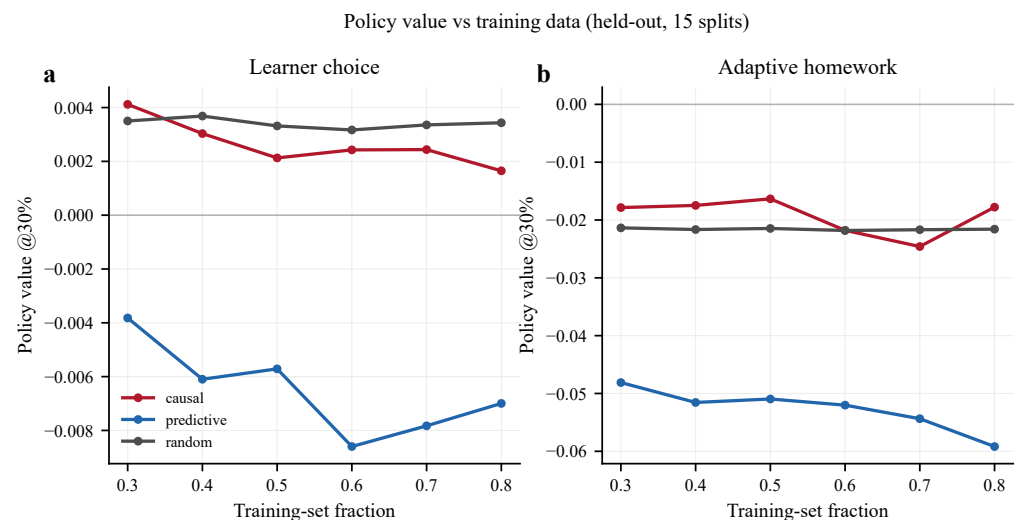


Figure 14. Policy value at 30% budget vs. training-set fraction (held-out, 15 splits). (a) Learner choice: the causal policy stays above the predictive policy at every training-set fraction, so the causal-over-predictive ordering is not an artifact of sample size. (b) Adaptive homework: the causal policy stays near the random baseline at every training-set fraction, so more training data do not create an advantage where none exists.

7. Discussion

Our results carry a concrete design implication for platforms that personalize learning-path features. The prevalent heuristic—deliver a feature to students predicted to perform well—optimizes the wrong quantity. Performance prediction identifies students who will succeed with or without the feature; it does not identify the students for whom the feature makes a difference. When a feature disproportionately helps weaker students (as learner choice does here), predictive targeting inverts the intended allocation and can do worse than random. Causal targeting, which ranks by the estimated benefit $\hat{\tau}$, corrects this inversion. Theorem 1 explains why (policy value depends only on the captured CATE), and Corollary 1 explains precisely when predictive targeting falls below random; related evaluations of targeting policies reach compatible conclusions [28].

Two practical guidelines follow. First, whether a feature should be personalized by student type is itself an empirical question with a clean test: estimate the CATE and evaluate its targeting quality on held-out data using the leakage-free statistic of Theorem 2. Where student-targetable heterogeneity exists (learner choice), causal targeting corrects the mis-allocation of predictive targeting; where it does not (adaptive homework), we find no evidence that the evaluated student-level targeting rules (causal, predictive, anti) beat simple non-personalized deployment, and the honest conclusion is not to personalize that feature by student type. Second, when heterogeneity is present, Theorem 3 gives a single deployable rule: rank once by $\hat{\tau}$ and recommend to those with $\hat{\tau} > 0$ (or to the top of the ranking under a budget).

We stress the modesty of the effect sizes. On standardized outcomes, the causal-versus-random gap is small and often not significant; the value of the method is not a large average lift but the correction of a systematic mis-allocation that a widely used heuristic introduces. In education, where a mis-targeted support feature can widen rather than narrow gaps, avoiding that inversion is itself the contribution.

7.1. Cost, Complexity, and Deployment Trade-Offs

Our expanded baseline suite (Section 6.7) tempers any enthusiasm for model complexity. A learned CATE model requires cross-fitted nuisance estimation, a benefit regressor, and split-sample evaluation; yet in this corpus it does not outperform a one-line risk-based rule (target low performers) or simply treating everyone. The reason is statistical rather than accidental. In this corpus, the feature's benefit is strongly and inversely aligned with baseline performance—the students predicted to perform worst are precisely those who gain most from learner choice (Section 6.17)—so the one-line risk proxy (target low $\hat{\mu}_0$) already recovers most of the benefit ordering that a CATE ranking would produce, leaving little marginal signal for a learned $\hat{\tau}$ to add. On the estimation side, with modest standardized effect sizes and only a few pre-treatment covariates, the learned $\hat{\tau}$ ranking carries estimation noise that can erase whatever small ordering advantage remains over a stable, low-variance risk proxy. By the same logic, a learned CATE model earns its complexity when these conditions fail: when benefit and baseline risk decouple (non-monotone or rank-crossing heterogeneity, so that low performance is no longer a good proxy for high benefit), when effect and sample sizes are large enough for the $\hat{\tau}$ ranking to be estimated precisely, or when richer covariates allow $\hat{\tau}$ to be estimated with low variance.

Quantified cost. A reviewer asked for concrete numbers rather than the assurance that training is cheap, and the request is fair: “trains in seconds” is not a deployment budget. Table 15 profiles the full pipeline on a single core of a commodity workstation. Training is indeed inexpensive—one end-to-end split costs 3.5 s (learner choice) or 1.6 s (adaptive homework), and the entire 40-split protocol that produces every point estimate in this paper costs 138 s and 65 s respectively. The deployed artifact is small: 700 trees, ~15,000 nodes, under 1 MB serialized, of which the benefit model $\hat{\tau}$ —the only component needed at serving time—is about 0.4 MB. Serving is likewise cheap: median single-student latency is 2.4 ms, the 99th percentile is 4.1 ms, and batch scoring reaches $\sim 4.5 \times 10^5$ students per second, so a platform-wide nightly re-rank of a million learners takes a few seconds. Peak memory stays near 210 MB.

Table 15. Measured computational cost, single core, per family. Training and serving are cheap; the one substantial cost is the offline permutation null, which refits the whole protocol 8000 times.

Quantity	Learner Choice	Adaptive Homework
Students/experiments	12,789/9	7708/8
Trees ($\hat{\mu}_1 200 + \hat{\mu}_0 200 + \hat{\tau} 300$)	700	700
Total nodes	15,344	13,776
Serialized model size (all components)	991.3 KB	906.2 KB
of which $\hat{\tau}$ alone (needed at serving)	438.2 KB	408.3 KB
Stage 1: cross-fitted nuisances	2.705 s	0.828 s
Stage 2: benefit model	0.703 s	0.749 s
Stage 3: policy + DR evaluation	0.049 s	0.038 s
One split, end to end	3.458 s	1.615 s
Full 40-split protocol	138.3 s	64.6 s
Single-student inference (median/p99)	2384.8/4102.7 μ s	2313.8/4088.7 μ s
Batch throughput	466,199 students/s	434,792 students/s
Peak working set	207.6 MB	212.0 MB
Permutation null (8000 refits), single-core	≈ 7.7 h	≈ 3.6 h
same, across 20 worker processes	≈ 23 min	≈ 11 min

The expensive part is inference, not training. We flag this explicitly because it is the one figure that is not negligible and that a naive cost profile would miss. The permutation null of Section 6.5 refits the entire pipeline under each permuted assignment: 200 replicates $\times$ 40 splits = 8000 full fits per family, which at the measured per-split cost is roughly 7.7 single-core hours for learner choice and 3.6 for adaptive homework. We ran it across 20 worker processes, which brings wall-clock time to well under an hour. This cost is paid once, offline, when deciding whether a feature should be personalized at all—never in the serving path.

Integration into a live platform. The pipeline needs three things from the host system and nothing else. (i) A period of randomized exposure—the feature toggled on/off at random for some students. This is the only genuinely intrusive requirement, and it is also what most A/B-testing infrastructure already provides; without it, the benefit model is not identified. (ii) A pre-treatment covariate snapshot taken strictly before the toggle, which rules out any feature computed from post-exposure activity. (iii) An outcome column. Given those, the fit is a nightly batch job of the order of minutes, and deployment is a lookup: score the student, compare $\hat{\tau}$ against zero (or against the budget quantile), show or withhold the feature. Because the serving artifact is a single ~ 0.4 MB tree ensemble with no external dependencies, it can be embedded directly in the recommendation service. We would nonetheless not recommend running it as a closed loop: once the policy governs exposure, the randomization that identified it is gone, so a persistent randomized holdout is necessary both to re-estimate benefit as the platform changes and to detect drift. Training and serving are inexpensive at this scale (Table 15); the one substantial computational cost is the offline permutation calibration. Compute is therefore not what should decide the question; the concern is value for complexity and operational risk. A CATE recommender adds moving parts that can drift, must be re-estimated per feature, and can silently mis-rank when covariates are weak—whereas risk-based targeting is transparent and auditable. Our practical recommendation is therefore staged: (i) never deploy predictive targeting, which is both value-negative and regressive; (ii) default to a transparent risk-based or treat-all policy; and (iii) invest in a learned CATE recommender only where a leakage-free test (Theorem 2) shows targetable heterogeneity and the estimated policy beats risk-based targeting out of sample. This inverts the common assumption that more sophisticated personalization is always better.

7.2. What a Practitioner Should Actually Do

The results above are a set of statistical findings; a reviewer asked what an instructional designer or platform team is supposed to do with them. We set out the sequence we would follow, with the decision rule at each step, and we are explicit about which steps our evidence supports and which are reasoned extrapolation.

Step 1: stop ranking by predicted performance. This is the one recommendation the evidence supports without qualification and the one that costs nothing to adopt. Delivering a support feature to the students a model predicts will do well answers the wrong question—it identifies students who will succeed with or without the feature. In our corpus, that rule attains a policy value of -0.074 at the 30% budget in adaptive homework, worse than random, and it excludes low-ability students at a rate 0.33 higher than it excludes high-ability ones. If a platform changes nothing else, it should stop doing this.

Step 2: ask whether the feature needs personalizing at all. Run one randomized on/off exposure and estimate the average effect. If the effect is uniformly positive and there is no evidence of student-level targetable heterogeneity—which is exactly what our DR-AUTO test with a permutation null detects (Theorem 2)—then the right policy is to give the feature to everyone, and no recommender is warranted. Adaptive homework in our corpus is the instructive case: its AUTO is -0.017 ($p = 0.69$), not significant, so personalizing it by student is not supportable, even though the family shows a large negative average effect that a designer would certainly want to investigate.

Step 3: if heterogeneity is present, check whether a simple rule already captures it. Before building a CATE model, plot mean doubly robust benefit against predicted baseline performance in deciles (Section 6.12). If the relationship is monotone—as in our learner-choice family, where benefit falls from $+0.087$ in the lowest decile to -0.030 in the highest—then “target the students predicted to struggle” recovers most of the achievable ordering, is transparent, is auditable by a teacher, and cannot silently mis-rank. In our corpus, this simple rule is at least as good as the learned CATE model at every budget. A learned model earns its complexity only when that relationship is non-monotone or rank-crossing, and our semi-synthetic sweep locates the crossover concretely (Section 6.12).

Step 4: if a CATE model is warranted, validate it on a split sample before trusting the ranking. Two diagnostics are cheap and decisive. Refit on two independent halves and correlate the two rankings: in our data, that correlation is ≈ 0.002 – 0.008 , which told us the individual-level ranking is near-noise even though the aggregate policy contrast is real. Then regress realized benefit on $\hat{\tau}$; a calibration slope far below 1 (ours is 0.018) means the model orders better than it measures, so it may be used to rank under a budget but not to make claims about how much any individual student will gain.

Step 5: set the budget from capacity, then read the value off the frontier. Because greedy top- $\hat{\tau}$ targeting traces the optimal frontier (Theorem 3), the budget is an operational input, not a modeling choice: pick b from licence seats, teacher time, or server cost, and the policy follows. When there is no binding constraint, the deployment rule is simply “recommend when $\hat{\tau} > 0$.”

We flag the boundary of this guidance honestly. Steps 1–2 are directly supported by the results in this paper. Steps 3–5 are supported as procedure—each diagnostic is one we ran and reported—but the thresholds a practitioner should use are not something a single corpus can calibrate, and the sign of the benefit–performance relationship in Step 3 is the platform-specific quantity we caution about in Section 8.

7.3. Fairness-Aware Deployment

A reviewer asked how the targeting policy should be governed once fairness is a stated objective, and our own results (Section 6.8) make clear why the question cannot

be waved away: the causal policy is markedly more equitable than predictive targeting on access, yet we cannot demonstrate that this converts into benefit for any particular group—every group-conditional policy value we estimate is indistinguishable from zero once dependence across students and experiments is accounted for. Access equity and benefit equity are different things, and optimizing one does not deliver the other.

Three concrete implications follow. First, audit at the group level, not only the aggregate. A policy with positive overall value can be harming an identifiable subgroup, and an aggregate number cannot reveal this. The audit we report—group-specific doubly robust policy value with intervals for every candidate policy, Table 9—is inexpensive once the DR scores exist and should be a standing part of deployment rather than a paper artifact. Our own audit is instructive precisely because it comes back inconclusive: at $n \approx 13,000$ students in nine experiments, the group-conditional intervals are too wide to certify any group as helped or harmed. A platform that runs this audit and reads a tight interval off non-independent splits will believe it has an answer when it does not. Second, treat an unresolved group value as a deployment blocker, not a footnote. The natural remedy is a constrained policy: maximize aggregate value subject to no group's value falling below zero (or below its treat-all value). We have not implemented or evaluated a constrained solver here, and we do not want to imply otherwise; that is future work. The prerequisite is a sample large enough to make the constraint testable, which this corpus is not. Third, do not assume the equitable policy is the sophisticated one. In our corpus, the one-line risk-based rule is more equitable on access (+0.328 low-minus-high gap versus +0.096 for causal targeting), and its point estimate is positive in every tercile, whereas the learned policy's is not. Where a simple rule is at least as good on both value and access, the accountable choice is the simple rule.

Finally, a governance point that our design forces. Whatever policy is deployed, it should be deployed against a persistent randomized holdout. Once a policy determines who sees a feature, the exposure is no longer random, group-specific benefit can no longer be estimated, and a policy that becomes inequitable as the population shifts will not be detectable. The holdout is the price of being able to keep answering the fairness question at all.

7.4. Applicability to Generative-AI Tutors

The method is agnostic to how a feature is implemented: it needs only randomized on/off exposure and an outcome, so it applies in principle to generative-AI tutoring features (LLM hints, dialogic tutors, adaptive feedback). Two cautions temper a direct transfer. First, the direction of heterogeneity need not carry over: whereas learner choice here helps academically weaker students, a recent randomized study of LLM tutoring finds the opposite—higher-prior-knowledge students gained more [43]—and other trials report both benefits and harms to retention [44,45]. This reinforces our central message that who benefits must be re-estimated per feature and never assumed, and that a mis-specified predictive rule could be even more regressive for gen-AI features. Second, and more limiting, we could not locate an openly released, participant-level, feature-randomized generative-AI tutoring dataset suitable for a clean policy-value replication; dialogue corpora exist but lack the randomized on/off contrast our estimator requires. A validation of causal targeting on generative-AI tutoring RCTs is therefore important future work rather than something we can demonstrate here.

8. Limitations

Several limitations bound our claims. (i) Setting. The evidence comes from a single platform (ASSISTments), one subject (middle-school mathematics), and one period

(2013–2015); generalization to other platforms, subjects, ages, or feature implementations is a matter for future work, not something our data can establish. We want to be concrete about what this restricts, because a reviewer rightly pressed on it. Three things are platform-specific rather than universal. First, the sign of the association between predicted performance and benefit: here, benefit falls with predicted performance (Section 6.10), which is what makes predictive targeting actively harmful. On a platform where a feature requires prerequisite fluency to exploit—a programming sandbox, an open-ended project—that association could invert, and predictive targeting could then be the better heuristic. Nothing in our data adjudicates this; the diagnostic in Section 6.12 is what a practitioner should run to find out. Second, the magnitude: standardized effects in short embedded experiments are small by construction, and the same features embedded in a semester-long curriculum could behave differently. Third, the covariate set: our five pre-treatment features are what this corpus records. A platform with richer telemetry (affect detectors, session-level engagement, teacher assignment) might support a materially better benefit model, and the near-noise ranking quality we document (Section 6.11) may be a property of this covariate set rather than of the method. What we believe does transfer is not a number but the framework and the failure mode: rank by benefit, not by predicted performance; evaluate on a split sample; and check whether the two rankings disagree before paying for the causal machinery. (ii) Magnitude and baselines. The contrast between causal and predictive targeting is significant in the adaptive-homework family and inconclusive in the learner-choice family once the dependence between splits is respected (Section 6.5); against a fuller baseline suite, the causal (CATE) policy does not beat random, a simple risk-based rule, or treat-all. So we do not claim CATE targeting is the best available policy—only that predictive targeting is the one to avoid. (iii) Estimated, not oracle, CATE. The theoretical optimality statements (Theorems 1 and 3) concern the population/oracle policy; our experiments use estimated $\hat{\tau}$ on finite held-out samples and therefore illustrate rather than prove those statements, and Section 6.11 quantifies how wide that gap is here: the estimated policy captures under 2% of the oracle gain, and its student-level ranking is close to irreproducible across independent halves. (iv) Construct scope. “Features” here are adaptive problem sequencing and a learner-choice interface; we do not study generative-AI tutoring, and the two families are not exhaustive. (v) Proxy and outcome. Outcomes are standardized within experiment, so effects are unitless and modest by construction; and the CATE is defined with respect to pre-randomization covariates that, while rich, do not capture unobserved motivation or context. (vi) Inference resolution. Our permutation null uses 200 replicates, so the smallest reportable two-sided p -value is $1/201 \approx 0.005$; “ $p \leq 0.005$ ” in this paper means exactly that and must not be read as $p < 0.001$. Raising the resolution is possible but costly—each replicate refits the full 40-split protocol—and would not change any conclusion, since no claim here turns on the difference between 0.005 and 0.001. Finally, the corpus and CATE machinery are shared with a companion empirical study; the contribution here is the targeting-policy method, its guarantees, and the causal-versus-predictive evaluation, not new claims about which students benefit.

9. Conclusions

We framed the recommendation of personalized learning-path features as a causal targeting-policy problem and gave it a theoretical and empirical treatment. Theoretically, causal top-CATE targeting is policy-value optimal at any budget and weakly dominates predictive targeting—strictly when their rankings disagree, and predictive targeting can even fall below random; a split-sample doubly robust evaluation of targeting quality is leakage-free while its naive in-sample counterpart is optimistically biased; and greedy top-CATE targeting traces the optimal cost-benefit frontier with the deployment rule

“recommend when $\hat{\tau} > 0$.” Empirically, on 17 randomized ASSISTments experiments and across 40 held-out splits, causal targeting outperforms predictive targeting in both feature families, and the contrast is statistically significant in the adaptive-homework family under a permutation null that respects the dependence between splits ($p \leq 0.005$; Holm-corrected $p \leq 0.040$ across the six primary tests). In the learner-choice family, the same contrast is directionally consistent at every budget and sign-stable under leave-one-experiment-out, but it does not reach significance under any of the three dependence-respecting tests we ran ($p = 0.23$ – 0.42), and we report it as inconclusive rather than confirmed. The mis-allocation being corrected is concrete: in the learner-choice family, benefit falls monotonically with predicted performance, so ranking students by predicted performance sends the feature to the students who benefit least. The practical message is not that causal targeting delivers a large lift over random—it does not, here, and a simple risk-based rule is at least as good—but that it removes a systematic, regressive error introduced by predicting performance instead of benefit and supplies a principled test for whether a feature should be personalized at all.

Author Contributions: Conceptualization, L.W.; Methodology, Y.H. and L.W.; Validation, L.W. and X.W.; Formal analysis, Y.H.; Investigation, Y.H. and W.B.; Resources, Y.H. and X.W.; Data curation, W.B. and X.W.; Writing—original draft, Y.H.; Writing—review & editing, W.B. and T.Z.; Visualization, W.B. and X.W.; Supervision, W.B. and T.Z.; Project administration, T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Guangdong Province Education Science Planning Project (2025GXJK0577, 2025GXJK0576), the Innovation Research Team Project of Zhaoqing University, the University-Level High-Caliber Training Project of Zhaoqing University (gcc202607), and the Funding Project for the ESI Discipline Advancement Plan at Zhaoqing University.

Data Availability Statement: The ASSISTments embedded-experiment corpus is publicly available from Prihar et al. [39]. All analysis code—data preparation and the leakage-safe contrast manifest, doubly robust nuisance and CATE estimation, the split-sample policy-value evaluator with the 40-repeat protocol, the permutation-based DR-AUTOC heterogeneity test, the expanded baseline suite, and every figure and table script—is released in the project repository for full reproducibility. Each experiment (E1–E10) writes a JSON results file consumed deterministically by the figure and table generators, so every reported number traces to a script and a raw output.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gao, C.; Zheng, Y.; Wang, W.; Feng, F.; He, X.; Li, Y. Causal inference in recommender systems: A survey and future directions. *ACM Trans. Inf. Syst.* **2024**, *42*, 88.
2. Wei, P.; Shu, H.; Gan, J.; Deng, X.; Liu, Y.; Sun, W.; Chen, T.; Hu, C.; Hu, Z.; Deng, Y.; et al. Sequential recommendation system based on deep learning: A survey. *Electronics* **2025**, *14*, 2134.
3. Lin, Y.; Wu, Z. Learning Path Recommendation Enhanced by Knowledge Tracing and Large Language Model. *Electronics* **2025**, *14*, 4385.
4. Athey, S.; Imbens, G. Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7353–7360.
5. Wager, S.; Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **2018**, *113*, 1228–1242.
6. Künzel, S.R.; Sekhon, J.S.; Bickel, P.J.; Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4156–4165.
7. Curth, A.; van der Schaar, M. Nonparametric Estimation of Heterogeneous Treatment Effects: From Theory to Learning Algorithms. *arXiv* **2021**, arXiv:2101.10943.
8. Nie, X.; Wager, S. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **2021**, *108*, 299–319.
9. Kennedy, E.H. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **2023**, *17*, 3008–3049.

10. Semenova, V.; Chernozhukov, V. Debiased machine learning of conditional average treatment effects and other causal functions. *Econom. J.* **2021**, *24*, 264–289.
11. Foster, D.J.; Syrgkanis, V. Orthogonal statistical learning. *Ann. Stat.* **2023**, *51*, 879–908.
12. Morzywolek, P.; Decruyenaere, J.; Vansteelandt, S. On Weighted Orthogonal Learners for Heterogeneous Treatment Effects. *arXiv* **2023**, arXiv:2303.12687.
13. Fan, Q.; Hsu, Y.C.; Lieli, R.P.; Zhang, Y. Estimation of conditional average treatment effects with high-dimensional data. *J. Bus. Econ. Stat.* **2022**, *40*, 313–327.
14. Kennedy, E.H.; Balakrishnan, S.; Robins, J.M.; Wasserman, L. Minimax rates for heterogeneous causal effect estimation. *Ann. Stat.* **2024**, *52*, 793–816.
15. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **2018**, *21*, C1–C68.
16. Johansson, F.D.; Shalit, U.; Kallus, N.; Sontag, D. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *J. Mach. Learn. Res.* **2022**, *23*, 1–50.
17. Bica, I.; Alaa, A.M.; Lambert, C.; van der Schaar, M. From real-world patient data to individualized treatment effects using machine learning: Current and future methods to address underlying challenges. *Clin. Pharmacol. Ther.* **2021**, *109*, 87–100.
18. Wang, D.; Chen, C.; Di, C.; Shu, M. Exploring behavior patterns for next-poi recommendation via graph self-supervised learning. *Electronics* **2023**, *12*, 1939.
19. Manski, C.F. Statistical treatment rules for heterogeneous populations. *Econometrica* **2004**, *72*, 1221–1246.
20. Kitagawa, T.; Tetenov, A. Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* **2018**, *86*, 591–616.
21. Athey, S.; Wager, S. Policy learning with observational data. *Econometrica* **2021**, *89*, 133–161.
22. Zhou, Z.; Athey, S.; Wager, S. Offline multi-action policy learning: Generalization and optimization. *Oper. Res.* **2023**, *71*, 148–183.
23. Kallus, N. More efficient policy learning via optimal retargeting. *J. Am. Stat. Assoc.* **2021**, *116*, 646–658.
24. Rzepakowski, P.; Jaroszewicz, S. Decision trees for uplift modeling with single and multiple treatments. *Knowl. Inf. Syst.* **2012**, *32*, 303–327.
25. Zhang, W.; Li, J.; Liu, L. A unified survey of treatment effect heterogeneity modeling and uplift modeling. *ACM Comput. Surv.* **2021**, *54*, 162.
26. Olaya, D.; Coussement, K.; Verbeke, W. A survey and benchmarking study of multitreatment uplift modeling. *Data Min. Knowl. Discov.* **2020**, *34*, 273–308.
27. Devriendt, F.; Guns, T.; Verbeke, W. Learning to rank for uplift modeling. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 4888–4904.
28. Yadlowsky, S.; Fleming, S.; Shah, N.; Brunskill, E.; Wager, S. Evaluating treatment prioritization rules via rank-weighted average treatment effects. *arXiv* **2021**, arXiv:2111.07966.
29. Wu, P.; Li, H.; Deng, Y.; Hu, W.; Dai, Q.; Dong, Z.; Sun, J.; Zhang, R.; Zhou, X.H. On the Opportunity of Causal Learning in Recommendation Systems: Foundation, Estimation, Prediction and Challenges. *arXiv* **2022**, arXiv:2201.06716.
30. Xu, S.; Ji, J.; Li, Y.; Ge, Y.; Tan, J.; Zhang, Y. Causal inference for recommendation: Foundations, methods and applications. *arXiv* **2023**, arXiv:2301.04016.
31. Zhu, Y.; Ma, J.; Li, J. Causal inference in recommender systems: A survey of strategies for bias mitigation, explanation, and generalization. *arXiv* **2023**, arXiv:2301.00910.
32. Jesson, A.; Mindermann, S.; Gal, Y.; Shalit, U. Quantifying ignorance in individual-level causal-effect estimates under hidden confounding. *arXiv* **2021**, arXiv:2103.04850.
33. Alkhatlan, A.; Kalita, J. Intelligent Tutoring Systems: A Comprehensive Historical Survey with Recent Developments. *arXiv* **2019**, arXiv:1812.09628.
34. Shen, S.; Liu, Q.; Huang, Z.; Zheng, Y.; Yin, M.; Wang, M.; Chen, E. A Survey of Knowledge Tracing: Models, Variants, and Applications. *IEEE Trans. Learn. Technol.* **2024**, *17*, 1898–1919.
35. Salau, L.; Hamada, M.; Prasad, R.; Hassan, M.; Mahendran, A.; Watanobe, Y. State-of-the-Art Survey on Deep Learning-Based Recommender Systems for E-Learning. *Appl. Sci.* **2022**, *12*, 11996.
36. Niu, Y.; Lin, R.; Xue, H. Research on Learning Resource Recommendation Based on Knowledge Graph and Collaborative Filtering. *Appl. Sci.* **2023**, *13*, 10933.
37. Ma, Y.; Wang, L.; Zhang, J.; Liu, F.; Jiang, Q. A Personalized Learning Path Recommendation Method Incorporating Multi-Algorithm. *Appl. Sci.* **2023**, *13*, 5946.
38. Bradáč, V.; Smolka, P.; Kotyrbá, M.; Průdek, T. Design of an Intelligent Tutoring System to Create a Personalized Study Plan Using Expert Systems. *Appl. Sci.* **2022**, *12*, 6236.
39. Prihar, E.; Syed, M.; Ostrow, K.; Shaw, S.; Sales, A.; Heffernan, N. Exploring Common Trends in Online Educational Experiments. In *Proceedings of the 15th International Conference on Educational Data Mining (EDM)*; International Educational Data Mining Society: Worcester, MA, USA, 2022.

40. Leite, W.L.; Zhang, H.; Rana, S.; Hao, Y.; Hatch, A.D.; Kong, L.; Kuang, H. Do intelligent tutoring systems benefit K-12 students? A meta-analysis and evaluation of heterogeneity of treatment effects in the U.S. *arXiv* **2025**, arXiv:2511.04997.
41. Schmucker, R.; Pachapurkar, N.; Bala, S.; Shah, M.; Mitchell, T. Learning to Optimize Feedback for One Million Students: Insights from Multi-Armed and Contextual Bandits in Large-Scale Online Tutoring. *arXiv* **2025**, arXiv:2508.00270.
42. Vanacore, K.; Thomas, D.R.; Smith, D.; Groot, B.; Reich, J.; Kizilcec, R. A Causal Framework for Estimating Heterogeneous Effects of On-Demand Tutoring. *arXiv* **2026**, arXiv:2602.19296.
43. Lehmann, M.; Cornelius, P.B.; Sting, F.J. AI Meets the Classroom: When Do Large Language Models Harm Learning? *arXiv* **2024**, arXiv:2409.09047.
44. Wang, R.E.; Ribeiro, A.T.; Robinson, C.D.; Loeb, S.; Demszky, D. Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise. *arXiv* **2024**, arXiv:2410.03017.
45. Huang, J.; Wang, R.R.; Liu, J.H.; Xia, B.; Huang, Y.; Sun, R.; Xue, J.M.; Zou, J. A Meta-Analysis of LLM Effects on Students across Qualification, Socialisation, and Subjectification. *arXiv* **2025**, arXiv:2509.22725.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.