THE JOINT-DISTILLATION GAP: WHICH TEACHER CORRECTIONS FIT TOGETHER UNDER SHARED STUDENT UPDATES?

Anonymous authors
Paper under double-blind review

ABSTRACT

Why can a shared student learn less from several specialized teachers than from each teacher separately, even when routing is correct and domain exposure is matched? We study this joint-distillation gap through the updates needed to learn teacher corrections. The cost of adding a correction depends on what the student is already learning, so a correction favored in isolation may be a poor addition to the selected supervision. This observation motivates Jointly Feasible Mode Distillation (JFMD). It decomposes teacher–student residuals into modes and selects corrections by their reliability-adjusted value and the projected cost of adding them to the current set. Across three model families and eight benchmarks, JFMD improves aggregate transfer over the evaluated student baselines. Controlled training comparisons across candidate pairs and matched-count ablations provide empirical support for conditional cost, while held-out analyses connect calibration reliability to stability on new examples. The gains persist across the tested student sizes and under teacher-routing errors. Together, the results suggest that learning requirements shared across corrections provide a useful explanation for the gap and a guide for selecting supervision. Our code is available at https://anonymous.4open.science/r/JFMD-392B.

1 INTRODUCTION

As large language models become increasingly specialized, models trained for mathematical reasoning Lewkowycz et al. (2022), science question answering Xu et al. (2024), translation Hsieh et al. (2023), and summarization Wan et al. (2024a) develop distinct strengths. Consolidating these capabilities into a compact student could reduce the cost of deploying and maintaining multiple models. Multi-teacher knowledge distillation offers a natural route to this goal by training one student with supervision from different domain teachers Jin et al. (2026). Yet supervision that is useful when learned from each teacher separately may behave differently when all teacher corrections must be learned through the same student parameters.

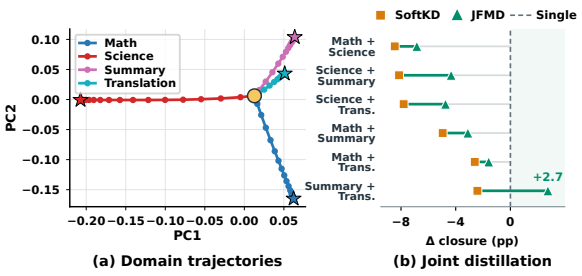


Figure 1: Joint-distillation gap. (a) Domain-specific trajectories from a shared initialization. (b) Teacher-gap closure of joint relative to separate SoftKD under oracle routing and matched exposure.

We examine this possibility by comparing joint and separate SoftKD from the same student initialization. Every training example is assigned to its corresponding domain teacher, and per-domain training exposure is matched between the two conditions. Even so, across six domain pairs, joint SoftKD closes 2.4–8.5 percentage points less of the normalized teacher–student performance gap than separate SoftKD (Figure 1b). Domain-specific adaptation trajectories also diverge in a shared PCA projection (Figure 1a). The trajectories illustrate that the domains follow different adaptation paths. The performance gap then raises the question of why learning them together yields less

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

transfer even when the teacher is correctly assigned and each domain receives the same training exposure.

The shared student parameters offer a clue because updates for different teacher corrections can affect one another. Prior work has studied related interactions through conflicting task gradients and cross-task update effects Yu et al. (2020); Fifty et al. (2021), and conflicts between teacher supervision signals have also been observed in multi-teacher distillation Du et al. (2020). Yet these interactions need not take the form of conflict. Two corrections may benefit from the same update, while others may demand competing updates. The difficulty of learning corrections together therefore depends on whether one student can accommodate their update requirements.

Across six domain pairs, greater additional update requirements for joint learning accompany larger losses relative to separate distillation. This association connects the performance gap to the cost of combining supervision and directs attention to how individual corrections are combined. We further find that the supervision already selected changes the additional cost of a new correction, altering candidate rankings relative to standalone costs. These findings motivate Jointly Feasible Mode Distillation (JFMD), which decomposes teacher–student corrections into residual modes and selects them by their reliability-adjusted value and additional cost given the modes already selected. The retained modes filter teacher corrections during distillation.

Across eight benchmarks, four domains, and three teacher–student model families, JFMD achieves the highest average discriminative and generative scores among the evaluated student methods in each family. With the number of retained modes matched, selection by additional cost improves transfer over selection by standalone cost. Across the evaluated candidate pairs, candidates preferred by conditional cost require less realized update in aggregate at matched progress while satisfying the preservation criterion for previously selected corrections. Held-out analyses show that calibration reliability tracks the stability of correction patterns on new examples beyond residual energy. We further find that JFMD retains its gains across the tested Qwen3 student sizes and is more robust to teacher-routing errors. Our contributions are summarized below.

- We document a joint-distillation gap across six domain pairs. Even with correct teacher assignments and matched per-domain training exposure, a shared student closes less of the teacher–student gap than students trained on the domains separately.
- We examine this gap through the update requirements of combined supervision. Across domain pairs, greater additional update requirements accompany larger joint-distillation losses. The cost of adding a correction also changes with the supervision already selected, reordering candidates relative to standalone costs.
- We develop JFMD to select and filter teacher corrections for a shared student. It combines reliability-adjusted residual-mode value with a projected cost conditional on the selected set. Across three model families and eight benchmarks, JFMD improves aggregate transfer over the evaluated student baselines, with gains across the tested student sizes and under teacher-routing errors.

2 Related Work

Multi-Teacher Knowledge Distillation. Multi-teacher knowledge distillation trains a shared student to learn from teachers with complementary expertise. Prior work transfers generative distributions from heterogeneous language models into a single student Wan et al. (2024b). Recent studies extend this approach to on-policy distillation, using student-generated rollouts to integrate capabilities from domain-specific teachers trained through reinforcement learning Ma et al. (2026). Another approach uses the strongest intermediate checkpoints for different domains as teachers to recover performance lost during successive stages of reinforcement learning Yang et al. (2026).

Task Interactions and Distillation Conflicts. Learning multiple tasks with shared parameters introduces interactions among their updates. Prior work mitigates gradient conflicts through projection Yu et al. (2020), accounts for worst-case local task improvement Liu et al. (2021), and formulates gradient combination as a bargaining game Navon et al. (2022). Cross-task update effects also help identify beneficial task groupings Fifty et al. (2021). In multi-teacher distillation, recent work identifies interference between recovering general capabilities and preserving domain expertise when

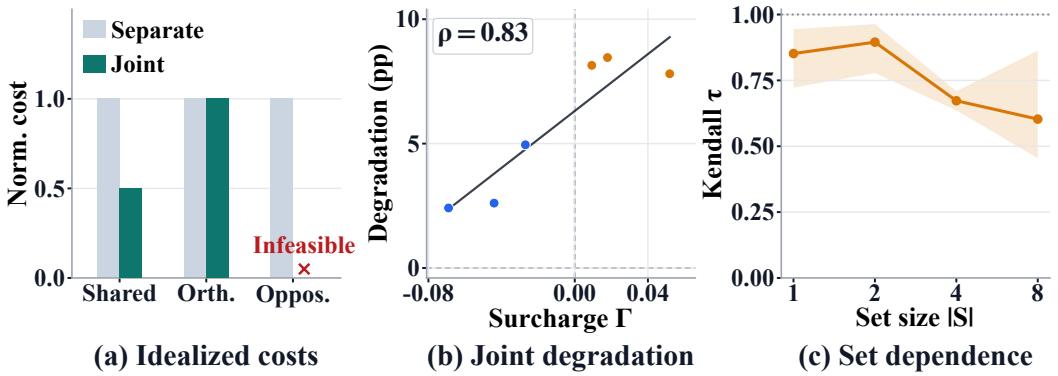


Figure 2: Joint update requirements and set-dependent cost. (a) Joint costs in controlled configurations, normalized by the sum of separate costs. (b) Feasibility surcharge versus joint-distillation degradation across six domain pairs. (c) Agreement between marginal-cost and singleton-cost rankings as the selected set grows.

teacher-aligned prompt coverage is incomplete Chen et al. (2026). Alternating recovery and preservation training addresses this interference, while gap-based sample selection prioritizes examples with greater correction needs .

3 Understanding the Joint-Distillation Gap

Figure 1 shows that a shared student gains less from the same domain supervision when several domains are learned together. We examine whether the teacher corrections require updates that can be shared by one student. Our units of analysis are residual modes, directions in the teacher–student residual over calibration contexts. Their construction is given in Section 4; here we use them to compare the updates required by separate corrections and by their combination.

3.1 Joint Update Requirements

Let S be a set of correction modes. For mode i, let $\mathcal{J}_i(\theta)$ measure the student's response, with gradient $g_i = \nabla_\theta \mathcal{J}_i(\theta_0)$ at the shared initialization and desired progress $\delta_i > 0$. Under a local approximation, an update $\Delta\theta$ supplies progress $g_i^\top \Delta\theta$. The minimum cost of an update that reaches every target in S is

$$C(S) = \min_{\Delta\theta} \frac{1}{2} \Delta\theta^\top M \Delta\theta \quad \text{s.t.} \quad g_i^\top \Delta\theta \geq \delta_i, \quad i \in S, \tag{1}$$

where $M \succ 0$ defines the update metric. We set $M = I$ to measure Euclidean update cost in LoRA parameter space. Section 4 specifies the response and progress target for each residual mode.

This cost depends on how the modes' requirements fit together. Modes with equal standalone costs can share an update when their requirements align, require separate progress when they are independent, or become jointly infeasible when their requirements oppose one another (Figure 2a). To examine this distinction across domains, let D be a set of domains and τ a common target for the fraction of teacher–student residual explained. Let $\rho_{\text{joint}}(D, \tau)$ be the minimum norm of one update that reaches this target in every domain in D, and $\rho_d(\tau)$ the minimum norm for domain d learned separately. We compare them through the joint-update surcharge

$$\Gamma(D, \tau) = \log \frac{\rho_{\text{joint}}(D, \tau)}{\sqrt{\sum_{d \in D} \rho_d(\tau)^2}}. \tag{2}$$

The denominator combines the separate-update norms in quadrature. A negative surcharge indicates update reuse relative to this reference; a positive one indicates that joint learning needs a larger update. Across six domain pairs, larger surcharges accompany greater losses in teacher-gap closure under joint relative to separate SoftKD (Figure 2b; Spearman $\rho = 0.83$). This exploratory association links the cost of combining supervision to the gap observed in training.

3.2 The Cost of Adding a Correction

The domain-pair association points to the cost of combining supervision. To use it when choosing what to teach, we must assess one candidate at a time. Once a set S of modes has been selected, candidate $i \notin S$ adds the following cost,

$$\Delta C_{i|S} = C(S \cup \{i\}) - C(S). \tag{3}$$

The update required by S may already advance i. Another candidate may need substantially more change to reach its target while maintaining progress on S. Thus, candidates that are inexpensive on their own need not remain inexpensive when added to the current supervision. Figure 2(c) compares candidate rankings by standalone and additional cost. Their mean Kendall agreement falls from about 0.90 to 0.60 as the selected set grows from two to eight modes.

We examine whether differences in candidate rankings correspond to realized learning difficulty through controlled training comparisons across multiple candidate pairs. For each pair, standalone cost favors candidate i, while additional cost given the selected set S favors candidate j. Both branches start from the same student adapted to S. For each candidate k, let R_k be the realized update required to reach a target progress while keeping degradation on S within a fixed tolerance, relative to continued training on S alone. The full metric and protocol are provided in Appendix B.2. Table 1 summarizes the results across the evaluated pairs. At the three matched progress targets, the aggregate update requirements are 10.9%, 42.9%, and 34.4% lower for candidates preferred by conditional cost than for those preferred by standalone cost. This aggregate ordering is consistent across the five seeds. These comparisons provide empirical support for evaluating a correction's learning cost relative to the selected supervision in the settings studied. The next section uses this quantity to guide the selection of teacher corrections for distillation.

Table 1: Aggregate update costs across candidate pairs at matched progress. R_i and R_j summarize the update requirements for candidates preferred by singleton and conditional cost, respectively. Savings are relative to R_i.

Progress	R_i	R_j	Savings
0.4	0.622	**0.554**	10.9%
0.5	1.873	**1.069**	42.9%
0.6	4.634	**3.039**	34.4%

4 Jointly Feasible Mode Distillation

The preceding analysis motivates selecting corrections by the value they carry and the additional cost of learning them alongside selected supervision. JFMD decomposes teacher corrections into modes, estimates their transferable value and set-conditional cost, selects modes under a cost budget, and filters teacher supervision through the selected modes (Figure 3).

Residual Modes and Transferable Value. For domain d, let $\mathcal{C}_d = \{x_t\}_{t=1}^{N_d}$ be a set of calibration contexts and let $\mathcal{V}_d$ be an anchor-vocabulary subset shared by the teacher and student. We define the teacher–student residual as

$$r_{d,t}(\theta) = (\log q_d(\cdot \mid x_t) - \log p_\theta(\cdot \mid x_t))_{\mathcal{V}_d}. \tag{4}$$

We stack the initial residuals $r_{d,t}(\theta_0)$ across calibration contexts and apply truncated SVD to obtain orthonormal residual modes $\{v_{d,k}\}_{k=1}^{K_d}$. Each mode represents a principal direction along which the teacher corrects the student. The calibration coefficient and energy of mode (d, k) are

$$a^{\mathrm{cal}}_{d,t,k} = r_{d,t}(\theta_0)^\top v_{d,k}, \qquad E^{\mathrm{cal}}_{d,k} = \frac{1}{N_d} \sum_t \left(a^{\mathrm{cal}}_{d,t,k}\right)^2. \tag{5}$$

A high-energy mode accounts for a large share of the calibration residual. To favor correction patterns that also persist across examples, we measure each full-bank mode's consistency across disjoint calibration splits. This gives a reliability score $\rho_{d,k} \in [0, 1]$. The transferable value of the mode is

$$\widehat{w}_{d,k} = \rho_{d,k} e_{d,k}, \qquad e_{d,k} = \frac{E^{\mathrm{cal}}_{d,k}}{\sum_{j=1}^{K_d} E^{\mathrm{cal}}_{d,j}}. \tag{6}$$

Thus a mode receives high value when it captures a substantial correction and its pattern remains consistent across calibration splits. The complete reliability definition is given in Appendix C.2.

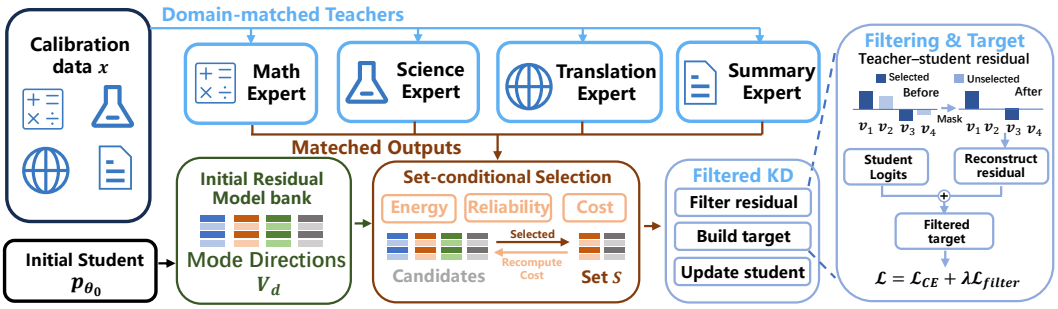


Figure 3: Overview of JFMD. Calibration decomposes teacher corrections into residual modes and estimates their transferable values and projected responses. Set-conditional selection retains modes under domain-coverage and update-cost constraints. The selected modes then reconstruct teacher corrections during distillation.

Mode Response and Joint Cost. We next estimate what the student must change to learn each mode. For mode $i = (d, k)$, fixed probe contexts $\mathcal{P}_d$ measure the student's response along that correction direction. Define the student response and the initial residual coefficient as

$$s_i(x_t; \theta) = v_{d,k}^\top \left(\log p_\theta(\cdot \mid x_t)\right)_{\mathcal{V}_d}, \qquad a_{t,i}^{\text{probe}} = v_{d,k}^\top r_{d,t}(\theta_0). \tag{7}$$

The response objective and desired progress are

$$\mathcal{J}_i(\theta) = \frac{1}{|\mathcal{P}_d|} \sum_{x_t \in \mathcal{P}_d} a_{t,i}^{\text{probe}} s_i(x_t; \theta), \qquad \delta_i = \frac{1}{|\mathcal{P}_d|} \sum_{x_t \in \mathcal{P}_d} \left(a_{t,i}^{\text{probe}}\right)^2. \tag{8}$$

Here, $\mathcal{J}_i$ measures the student's response along mode i, while δ_i is the progress reached when that response matches the teacher on the probe contexts. The corresponding gradient $g_i = \nabla_\theta \mathcal{J}_i(\theta_0)$ describes how a local student update advances the mode. Together, g_i and δ_i instantiate the cost $C(S)$ from Section 3.1.

Conditional Cost and Budget. The value $\widehat{w}_i$ captures how much reliable correction a mode carries, while $C(S)$ captures the cost of learning a selected set together. Let w_i denote a candidate's assigned value and let B be an update-cost budget. The idealized selection problem is

$$S^\star \in \arg\max_{S \subseteq \mathcal{I}} \sum_{i \in S} w_i, \qquad \text{s.t.} \quad C(S) \le B. \tag{9}$$

For JFMD, $w_i = \widehat{w}_i$ from Equation 6. The idealized cost $C(S)$ requires every selected mode to reach its progress target. To evaluate candidate sets during selection, JFMD uses $\widehat{C}(S)$, a projected response-matching proxy that penalizes remaining mismatch.

With $M = I$, we project each response gradient into a shared space, $\widetilde{g}_i = P g_i$, where $P \in \mathbb{R}^{m \times p}$. For a set S, let $\widetilde{G}_S$ stack the projected gradients as rows and let δ_S collect their progress targets. Regularized response matching balances update size against squared target mismatch. Its optimal value is

$$\widehat{C}(S) = \frac{1}{2} \delta_S^\top \left(\widetilde{G}_S \widetilde{G}_S^\top + \eta I\right)^{-1} \delta_S, \qquad \eta > 0. \tag{10}$$

The projected Gram matrix retains interactions among the modes in this estimate. With $\widehat{C}(\varnothing) = 0$, the surrogate is monotone, so $\widehat{C}(S \cup \{i\}) \ge \widehat{C}(S)$. Its marginal increment estimates the additional cost of a candidate given the modes already selected. Derivations and the closed-form marginal update are provided in Appendix B.3.

Budgeted Mode Selection. We first retain modes that meet the reliability threshold, $\mathcal{I}_{\text{elig}} = \{i : \rho_i \geq r_{\min}\}$. Given a selected set S, we compare remaining modes by value per additional cost.

$$\text{Score}(i \mid S) = \frac{\widehat{w}_i}{\widehat{C}(S \cup \{i\}) - \widehat{C}(S) + \varepsilon}, \tag{11}$$

where $\varepsilon > 0$ ensures numerical stability. After ensuring domain coverage, JFMD adds feasible high-scoring candidates under a proxy-cost budget, recomputing scores after each addition. The effective budget is at least the cost of the coverage set. The coverage rule and stopping criterion are specified in Appendix B.4.

Mode-Filtered Distillation. Let $z_{d,k} \in \{0, 1\}$ indicate whether mode (d, k) is selected. During training, we project the current teacher–student residual onto the span of the selected modes:

$$\widehat{r}_{d,t}(\theta) = \sum_{k=1}^{K_d} z_{d,k} \left(r_{d,t}(\theta)^\top v_{d,k}\right) v_{d,k}. \tag{12}$$

This projection retains the selected correction directions. We add the reconstructed correction to the student logits to form a detached distillation target:

$$\widetilde{q}_{d,t} = \text{softmax}\big(\text{sg}\big[\ell_\theta(x_t) + \mathcal{E}_d\big(\widehat{r}_{d,t}(\theta)\big)\big]\big), \tag{13}$$

where $\ell_\theta(x_t)$ denotes the student logits, $\mathcal{E}_d$ embeds the anchor-vocabulary correction into the full vocabulary, and sg stops gradients. The student is trained with

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \left\langle \text{KL}\left(\widetilde{q}_{d,t} \,\|\, p_\theta(\cdot \mid x_t)\right)\right\rangle. \tag{14}$$

The selected basis remains fixed during training, while its residual coefficients are recomputed from the current student. As the student learns a selected correction, the residual along that mode changes accordingly. The sparse-support implementation is given in Appendix C.4.

5 Experiments

The preceding analysis leads to a test of the intervention. If the cost of learning a teacher correction depends on the supervision already selected, choosing corrections by their additional cost should improve joint distillation. We evaluate JFMD across three model families and eight benchmarks, then use matched-count comparisons to examine which choices produce the gains. We also ask whether the selection signals observed during calibration hold for new examples, and whether the gains persist as student capacity and teacher-routing quality change.

5.1 Experimental Setup

Datasets and evaluation metrics. We evaluate eight benchmarks across four domains: GSM8K Cobbe et al. (2021) and MATH Hendrycks et al. (2021) for mathematics, SciQ Welbl et al. (2017) and OpenBookQA (OBQA) Mihaylov et al. (2018) for science, FLORES Team et al. (2022) and IWSLT Cettolo et al. (2017) for translation, and XSum Narayan et al. (2018) and SAMSum Gliwa et al. (2019) for summarization. We report accuracy for mathematics and science, BLEU Papineni et al. (2002) for translation, and ROUGE-L Lin (2004) for summarization, along with separate averages over the four discriminative and four generative tasks.

Baselines and training settings. We evaluate Qwen3-8B→Qwen3-1.7B, Llama-3.1-8B→Llama-3.2-1B, and Gemma-7B→Gemma-2B Yang et al. (2025); Touvron et al. (2023); Gemma Team et al. (2024). Baselines include SFT, SoftKD Hinton et al. (2015), WeightedKD Du et al. (2020), MiniLLM Gu et al. (2024), GKD Agarwal et al. (2024), and AdaKD Xie et al. (2026). PCGrad Yu et al. (2020) and reference-KL regularization Ouyang et al. (2022) provide training-time controls. Within each model family, comparisons share the student initialization, domain teachers, source data, and training and validation split. Main experiments route examples by their known domain labels. SoftKD and JFMD use the same KD support and loss averaging; SoftKD uses the routed teacher target, while JFMD uses its filtered target. We report means over five seeds and standard deviations for aggregate metrics. JFMD's reported time includes calibration, mode construction, selection, and training. Further implementation details are in Appendix C.

Table 2: Multi-domain distillation from specialized teachers to shared compact students. average columns also report standard deviations. Bold and underlined values denote the best and second-best results among trained student methods within each model family. Time includes calibration and mode-bank construction for JFMD.

	Math		Science		Translation		Summarization		Average		Time
Method	GSM8K Acc.	MATH Acc.	SciQ Acc.	OBQA Acc.	FLORES BLEU	IWSLT BLEU	XSum R-L	SAMSum R-L	Acc. Avg.	Gen. Avg.	(h)
Qwen3: 8B → 1.7B											
Teachers	81.42	69.38	77.60	88.09	37.02	25.12	32.14	47.82	79.12	35.53	—
Student	61.43	42.22	64.88	70.00	25.27	15.48	12.21	29.35	59.63	20.58	—
SFT	67.65	62.71	65.31	72.00	25.85	15.85	27.07	45.27	66.92 ± 0.56	28.51 ± 0.24	2.08
SoftKD	68.02	62.99	65.52	72.52	26.85	14.72	27.22	45.50	67.26 ± 0.33	28.57 ± 0.38	2.96
WeightedKD	66.51	59.41	48.40	56.51	25.25	15.18	26.51	44.85	57.71 ± 1.49	27.95 ± 0.19	5.32
MiniLLM	56.48	56.91	**69.22**	76.48	25.03	17.43	12.22	28.97	64.77 ± 0.99	20.91 ± 0.21	15.42
GKD	71.02	**67.14**	61.94	67.42	21.77	17.29	27.13	45.55	66.88 ± 0.77	27.94 ± 0.30	20.34
AdaKD	71.01	66.09	66.46	71.93	23.60	17.33	26.89	**45.99**	68.87 ± 0.57	28.45 ± 0.14	21.12
JFMD	**72.86**	65.25	66.27	**76.67**	**27.52**	**17.73**	**28.20**	45.95	**70.26** ± 0.15	**29.85** ± 0.29	4.07
LLaMA: 8B → 1B											
Teachers	55.44	24.81	72.68	81.83	29.70	25.73	35.49	48.13	58.69	34.76	—
Student	2.88	3.38	20.40	20.66	0.17	0.28	9.27	13.03	11.83	5.69	—
SFT	12.61	6.13	57.01	43.95	14.18	13.97	27.42	43.17	29.93 ± 0.41	24.68 ± 0.13	2.01
SoftKD	13.18	6.11	57.92	46.16	14.84	14.43	28.10	43.36	30.84 ± 0.19	25.18 ± 0.22	2.68
WeightedKD	14.02	6.90	59.60	53.90	16.11	12.60	29.02	44.11	33.61 ± 0.08	25.46 ± 0.90	3.86
MiniLLM	13.02	6.54	56.73	49.55	13.95	13.69	28.13	43.34	31.46 ± 0.47	24.78 ± 0.16	10.42
GKD	14.21	6.32	54.73	52.36	14.15	13.87	29.04	43.71	31.91 ± 0.31	25.19 ± 0.21	12.04
AdaKD	13.47	7.03	56.76	51.83	14.86	13.41	28.80	43.26	32.27 ± 1.00	25.08 ± 0.29	12.68
JFMD	**15.05**	**7.31**	**60.87**	**55.18**	**17.97**	**14.91**	**29.66**	**44.73**	**34.60** ± 0.51	**26.82** ± 0.18	3.13
Gemma: 7B → 2B											
Teachers	74.96	51.09	80.96	90.35	38.95	30.27	37.03	50.52	74.34	39.19	—
Student	2.41	3.32	16.38	16.45	0.18	0.32	9.48	16.31	9.64	6.57	—
SFT	31.66	19.65	66.55	72.08	27.19	23.38	32.12	45.50	47.49 ± 0.29	32.05 ± 0.13	2.71
SoftKD	33.13	19.76	67.90	74.01	28.58	24.13	32.50	45.83	48.70 ± 0.21	32.76 ± 0.42	3.22
WeightedKD	32.51	19.70	67.69	74.05	28.02	24.14	32.56	46.09	48.49 ± 0.67	32.70 ± 0.02	5.94
MiniLLM	32.90	19.80	68.74	74.44	28.22	23.95	32.68	45.90	48.97 ± 0.45	32.69 ± 0.14	21.32
GKD	**33.51**	**22.46**	63.43	74.06	27.21	22.53	31.88	**46.82**	48.37 ± 0.55	32.11 ± 0.31	27.31
AdaKD	32.87	19.38	**69.26**	74.70	28.98	23.56	32.32	46.36	49.05 ± 0.45	32.81 ± 0.28	26.04
JFMD	33.40	21.96	68.87	**75.87**	**29.36**	**24.68**	**33.44**	46.68	**50.02** ± 0.23	**33.54** ± 0.17	4.33

5.2 Overall Performance and Efficiency

Under teacher assignments fixed by domain labels, we compare multi-domain transfer to a shared student. Table 2 shows that JFMD achieves the highest average accuracy and generation score among the evaluated student methods in each of the three model families. Relative to the strongest baseline for each aggregate metric, its accuracy gains are 1.39, 0.99, and 0.97 points on Qwen3, LLaMA, and Gemma, with generation gains of 1.28, 1.36, and 0.73 points, respectively. On Qwen3, GKD improves MATH over SoftKD but loses ground on SciQ, while MiniLLM shows the reverse. These crossed changes are consistent with teacher corrections interacting through shared student parameters. Progress on one domain can accompany weaker transfer on another, producing uneven task-level outcomes.

JFMD retains the SoftKD training loss and filters the teacher corrections used to form its distillation target. Its higher aggregate scores than SoftKD under the same teacher assignments show the benefit of this selection in the tested settings. The selection adds calibration and mode-construction time. Including these stages, JFMD takes 3.13–4.33 hours, compared with 2.68–3.22 hours for SoftKD, and runs faster than MiniLLM, GKD, and AdaKD in these configurations.

Table 3: Training controls and mode-selection ablations across three model families. Selection variants share the source mode bank and match JFMD's retained mode count in each domain. Bold and underlined values mark the best and second-best trained student methods within each family.

	Qwen3: 8B → 1.7B		LLaMA: 8B → 1B		Gemma: 7B → 2B	
Variant	Acc. Avg.	Gen. Avg.	Acc. Avg.	Gen. Avg.	Acc. Avg.	Gen. Avg.
SoftKD	67.26 ± 0.33	28.57 ± 0.38	30.84 ± 0.19	25.18 ± 0.22	48.70 ± 0.21	32.76 ± 0.42
w/ PCGrad	63.59 ± 0.44	28.34 ± 0.21	31.19 ± 0.65	25.26 ± 0.58	44.87 ± 0.38	31.93 ± 0.11
w/ Reference KL	68.05 ± 0.69	28.54 ± 0.59	32.60 ± 0.45	25.98 ± 0.10	48.03 ± 0.18	32.22 ± 0.02
JFMD	**70.26** ± 0.15	**29.85** ± 0.29	**34.60** ± 0.51	**26.82** ± 0.18	**50.02** ± 0.23	**33.54** ± 0.17
w/o Cost	65.01 ± 1.85	29.11 ± 0.23	34.30 ± 0.28	26.37 ± 0.38	49.19 ± 0.22	33.36 ± 0.18
w/o Reliability & Cost	63.36 ± 1.19	29.34 ± 0.50	<u>34.43</u> ± 0.23	26.60 ± 0.63	49.19 ± 0.22	<u>33.40</u> ± 0.15
w/ Random Selector	64.42 ± 1.18	29.39 ± 0.23	33.60 ± 0.52	<u>26.67</u> ± 0.30	48.49 ± 0.35	33.17 ± 0.35
w/ Separable Selector	68.03 ± 1.85	<u>29.43</u> ± 0.39	34.13 ± 0.40	26.35 ± 0.29	49.27 ± 0.22	33.22 ± 0.13
w/ Singleton-Cost	<u>68.51</u> ± 0.66	29.12 ± 0.32	33.88 ± 0.33	26.10 ± 0.13	<u>49.29</u> ± 0.17	32.99 ± 0.36

5.3 Effect of Supervision Selection

The overall comparison shows JFMD's aggregate gains across three model families. We next examine how the choice of teacher corrections relates to these gains. Selection variants share the source mode bank and downstream training settings, and retain the same number of modes per domain, without matching the cost budget. Removing cost leaves reliability-adjusted energy as the ranking criterion; removing reliability as well leaves energy alone. Random selection uses neither ranking. The w/ Separable Selector variant retains the reliability threshold and uses a sensitivity-adjusted per-mode cost, while w/ Singleton-Cost replaces marginal cost with each mode's standalone cost.

Under these matched conditions, JFMD has higher mean scores than random and energy-only selection on all six aggregate metrics (Table 3). On Qwen, average accuracy rises from 65.01 with value-only selection to 68.03 with the separable selector, 68.51 with singleton cost, and 70.26 with JFMD's set-conditional cost. Across the three families, JFMD also has higher mean scores than both cost controls on all six aggregate metrics. The singleton-to-conditional ordering is consistent with the aggregate results across candidate pairs in Section 3.2, where candidates preferred by conditional cost required less realized update at matched progress under the preservation criterion. PCGrad modifies competing gradients, while reference-KL constrains student deviation during training. Both controls have lower aggregate mean scores than JFMD in these configurations. Training-time controls alone do not match the aggregate gains observed with correction selection here.

5.4 From Conditional Cost to Stable Corrections

The preceding ablation favors evaluating each candidate relative to the supervision already selected. We next examine whether JFMD's implemented cost preserves this ordering. In the same projected space, rankings from its regularized cost closely follow those from the inequality-constrained cost, with five-seed mean Spearman correlations above 0.995 across the tested sets (Table 5). JFMD combines this cost with a mode's value when making each selection. That value includes reliability measured across calibration splits. Figure 4 shows that higher-reliability modes remain more stable on new examples, including among modes with similar calibration energy (Appendix A.5). The implemented selector thus considers both the cost of adding a correction to the current set and evidence that its pattern persists beyond calibration.

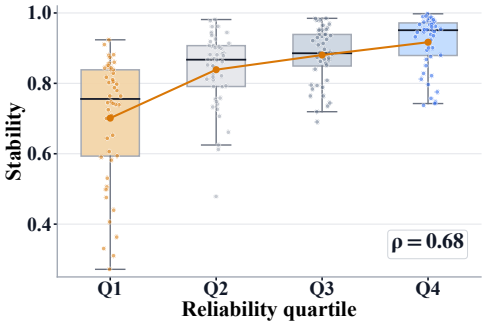


Figure 4: Held-out stability by reliability quartile ($\rho = 0.68$).

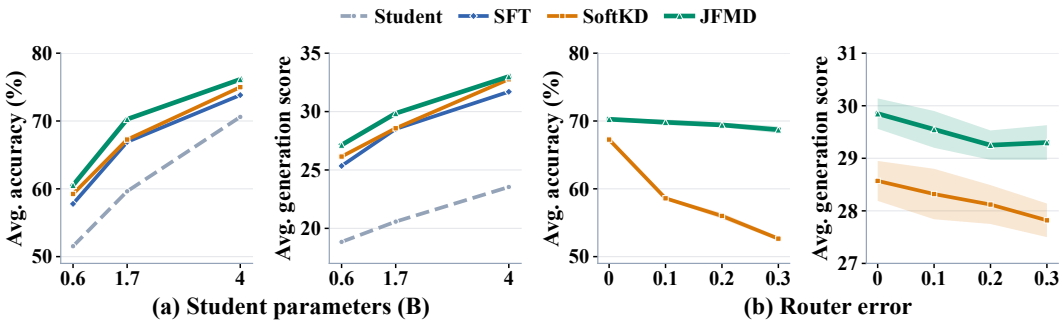


Figure 5: Performance across student sizes and teacher-routing error rates. (a) Average accuracy and generation score for Qwen3 students at three sizes. (b) SoftKD and JFMD as the routing-error rate increases.

5.5 Student Capacity and Routing Robustness

Student capacity. Larger students may accommodate more teacher corrections, reducing the need for selection. Yet JFMD outperforms SoftKD on both aggregate metrics at all three Qwen3 sizes (Figure 5(a)), with the largest gap at the intermediate scale. Selection gains therefore do not simply diminish as capacity increases, and choosing which corrections to retain remains useful across the tested sizes (Appendix D.2).

Routing errors. Beyond limited capacity, a shared student may receive corrections from an incorrectly assigned teacher. JFMD retains its advantage under this additional challenge. At a 30% routing-error rate, accuracy drops by 14.61 points for SoftKD but only 1.55 for JFMD (Figure 5(b)). Both methods degrade less on generation tasks, where JFMD maintains its lead. To understand what filtering changes under incorrect routing, we separately inject the retained and removed components of erroneous corrections. Relative to the correctly routed JFMD target, retained errors do not increase correct-answer NLL on average. Injecting only the removed component produces a larger NLL increase than injecting the full error, even though the removed component contains less energy (Figure 6(a)). This contrast extends to the resulting training signals. On average, Retained errors and correct routing yield positive KD–supervised gradient alignment, while removed and full errors yield negative alignment (Figure 6(b)). Together, these local diagnostics associate the excluded corrections with poorer targets and conflicting gradients, consistent with the observed routing robustness.

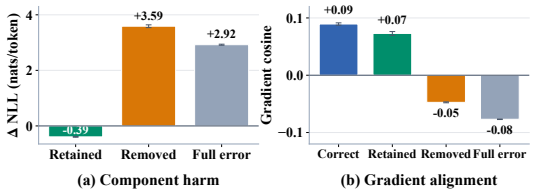


Figure 6: Effects of routing errors. (a) Target-distribution answer NLL change relative to the correctly routed JFMD target. (b) KD–supervised gradient similarity.

6 Conclusion

Joint distillation leaves a larger teacher–student gap than separate distillation even with correct routing and matched exposure. Our analyses associate this gap with interactions among teacher corrections through shared student updates. Guided by this observation, JFMD selects residual modes by reliability-adjusted value and their projected cost conditional on selected supervision. Across three model families and eight benchmarks, it achieves the highest aggregate scores among the evaluated student methods. Controlled training, matched-count ablations, and held-out analyses support its selection criteria, while gains persist across the tested student sizes and under teacher-routing errors. These results suggest that selected supervision provides a useful context for choosing further corrections in the settings studied here.

AI USE STATEMENT

During the preparation of this manuscript, we used AI tools to assist with language polishing and text editing, including improvements to grammar, wording, sentence structure, and clarity.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

Mauro Cettolo, Marcello Federico, Luisa Bentivogli, Jan Niehues, Sebastian Stüker, Katsuhito Sudoh, Koichiro Yoshino, and Christian Federmann. Overview of the IWSLT 2017 evaluation campaign. In Sakriani Sakti and Masao Utiyama (eds.), *Proceedings of the 14th International Conference on Spoken Language Translation*, pp. 2–14, Tokyo, Japan, December 14-15 2017. International Workshop on Spoken Language Translation. URL https://aclanthology.org/2017.iwslt-1.1/.

Tianlei Chen, Jiao Ou, Ziyuan Liu, Ruiming Tang, Jian Liang, and Han Li. Counteraction-aware multi-teacher on-policy distillation for general capability recovery with domain preservation. *arXiv preprint arXiv:2605.27115*, 2026.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL https://arxiv.org/abs/2110.14168.

Shangchen Du, Shan You, Xiaojie Li, Jianlong Wu, Fei Wang, Chen Qian, and Changshui Zhang. Agree to disagree: Adaptive ensemble knowledge distillation in gradient space. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 12345–12355. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/91c77393975889bd08f301c9e13a44b7-Paper.pdf.

Chris Fifty, Ehsan Amid, Zhe Zhao, Tianhe Yu, Rohan Anil, and Chelsea Finn. Efficiently identifying task groupings for multi-task learning. *Advances in Neural Information Processing Systems*, 34:27503–27516, 2021.

Gemma Team et al. Gemma 2: Improving open language models at a practical size, 2024. URL https://arxiv.org/abs/2408.00118.

Bogdan Gliwa, Iwona Mochol, Maciej Biesek, and Aleksander Wawer. SAMSum corpus: A human-annotated dialogue dataset for abstractive summarization. In Lu Wang, Jackie Chi Kit Cheung, Giuseppe Carenini, and Fei Liu (eds.), *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, pp. 70–79, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-5409. URL https://aclanthology.org/D19-5409/.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 32694–32717, 2024.

N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011. doi: 10.1137/090771806. URL https://doi.org/10.1137/090771806.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. In J. Vanschoren and S. Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/be83ab3ecd0db773eb2dc1b0a17836a1-Paper-round2.pdf.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015. URL https://arxiv.org/abs/1503.02531.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 8003–8017, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.507. URL https://aclanthology.org/2023.findings-acl.507/.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL https://arxiv.org/abs/2106.09685.

Ruihan Jin, Pengpeng Shao, Zhengqi Wen, Jinyang Wu, Mingkuan Feng, Shuo Yang, Chu Yuan Zhang, and Jianhua Tao. Exploring knowledge purification in multi-teacher knowledge distillation for llms. *arXiv preprint arXiv:2602.01064*, 2026.

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, et al. Solving quantitative reasoning problems with language models. *Advances in neural information processing systems*, 35:3843–3857, 2022.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013/.

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in neural information processing systems*, 34:18878–18890, 2021.

Wenhan Ma, Jianyu Wei, Liang Zhao, Hailin Zhang, Bangjun Xiao, Lei Li, Qibin Yang, Bofei Gao, Yudong Wang, Rang Li, et al. Mopd: Multi-teacher on-policy distillation for capability integration in llm post-training. *arXiv preprint arXiv:2606.30406*, 2026.

Todor Mihaylov, Peter Clark, Tushar Khot, and Ashish Sabharwal. Can a suit of armor conduct electricity? a new dataset for open book question answering. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2381–2391, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1260. URL https://aclanthology.org/D18-1260/.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1797–1807, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1206. URL https://aclanthology.org/D18-1206/.

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. *arXiv preprint arXiv:2202.01017*, 2022.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin (eds.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL https://aclanthology.org/P02-1040/.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No language left behind: Scaling human-centered machine translation, 2022. URL https://arxiv.org/abs/2207.04672.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023. URL https://arxiv.org/abs/2302.13971.

Fanqi Wan, Xinting Huang, Deng Cai, Xiaojun Quan, Wei BI, and Shuming Shi. Knowledge fusion of large language models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 18303–18322, 2024a. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/4fd5cfd2e31bebbccfa5ffa354c04bdc-Paper-Conference.pdf.

Fanqi Wan, Xinting Huang, Deng Cai, Xiaojun Quan, Wei Bi, and Shuming Shi. Knowledge fusion of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 18303–18322, 2024b.

Johannes Welbl, Nelson F. Liu, and Matt Gardner. Crowdsourcing multiple choice science questions. In Leon Derczynski, Wei Xu, Alan Ritter, and Tim Baldwin (eds.), *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pp. 94–106, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-4413. URL https://aclanthology.org/W17-4413/.

Xurong Xie, Zhucun Xue, Jiafu Wu, Jian Li, Yabiao Wang, Xiaobin Hu, Yong Liu, and Jiangning Zhang. Llm-oriented token-adaptive knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 34070–34078, 2026.

Haoran Xu, Young Jin Kim, Amr Mohamed Nabil Aly Aly Sharaf, and Hany Awadalla. A paradigm shift in machine translation: Boosting translation performance of large language models. In *International Conference on Learning Representations*, volume 2024, pp. 2747–2767, 2024.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. URL https://arxiv.org/abs/2505.09388.

Zhuolin Yang, Zihan Liu, Yang Chen, Wenliang Dai, Boxin Wang, Sheng-Chieh Lin, Chankyu Lee, Yangyi Chen, Dongfu Jiang, Jiafan He, et al. Nemotron-cascade 2: Post-training llms with cascade rl and multi-domain on-policy distillation. *arXiv preprint arXiv:2603.19220*, 2026.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

APPENDIX CONTENTS

A Low-Dimensional Structure of Supervision **14**

A.1 Local Response Mapping (Q1) 14

A.2 Sufficiency of the Response Subspace (Q2) 15

A.3 Projection and Cost Preservation (Q3) 16

A.4 Residual Reconstruction and Mode Energy (Q4) 17

A.5 Cross-Sample Residual Stability (Q5) 18

B Joint Cost and Set-Conditional Selection **20**

B.1 Minimum-Update Cost and Update Reuse (Q6) 20

B.2 Set-Dependent Candidate Ordering (Q7) 22

B.3 Tractable Joint Cost and Marginal Increments (Q8) 23

B.4 Supervision Selection with Domain Coverage (Q9) 25

C Implementation and Experimental Protocols **27**

C.1 Teacher–Student Configurations and Data 27

C.2 Calibration and Mode Construction 28

C.3 Response Estimation and Projection 29

C.4 Distillation Targets and Training 29

C.5 Baselines and Controlled Comparisons 30

C.6 Evaluation and Computational Cost 31

D Extended Results and Robustness **31**

D.1 Task-Level Results of Controlled Ablations 31

D.2 Transfer across Student Capacities 31

D.3 Robustness to Teacher Routing Errors 32

D.4 Sensitivity to Selection Settings 35

A Low-Dimensional Structure of Supervision

This section examines how teacher corrections and student responses can be represented in fewer dimensions, and what these representations preserve. The analysis follows five questions:

- Q1. How do parameter updates translate into progress on supervision components?
- Q2. Which parameter directions suffice to realize selected progress targets under the local response model?
- Q3. When does a practical projection preserve regularized costs and marginal-cost orderings?
- Q4. What correction information do residual modes and their energies retain?
- Q5. Under what conditions does calibrated residual structure persist across samples?

A.1 Local Response Mapping (Q1)

A shared student receives corrections in output space but learns through parameter updates. To connect the two, we fix the calibrated mode directions and describe how the student's response along each direction changes with its parameters.

Mode responses and target progress. For domain d, let $\mathcal{P}_d$ contain N_d^{probe} probe token contexts. For mode $i = (d, k)$, define its initial coefficient using the unclipped log-probability residual:

$$a_{t,i}^{\mathrm{probe}} = v_{d,k}^{\top} \left(\log q_d(\cdot \mid x_t) - \log p_{\theta_0}(\cdot \mid x_t)\right)_{\mathcal{V}_d}. \tag{15}$$

Holding these coefficients and directions fixed, the response objective is

$$\mathcal{J}_i(\theta) = \frac{1}{N_d^{\mathrm{probe}}} \sum_{x_t \in \mathcal{P}_d} a_{t,i}^{\mathrm{probe}} v_{d,k}^{\top} \big(\log p_\theta(\cdot \mid x_t)\big)_{\mathcal{V}_d}. \tag{16}$$

Let $\mathcal{J}_i^T$ denote the same weighted response evaluated with the teacher distribution. Then

$$\begin{aligned} \mathcal{J}_i^T - \mathcal{J}_i(\theta_0) &= \frac{1}{N_d^{\mathrm{probe}}} \sum_{x_t \in \mathcal{P}_d} a_{t,i}^{\mathrm{probe}} v_{d,k}^{\top} r_{d,t}(\theta_0) \\ &= \frac{1}{N_d^{\mathrm{probe}}} \sum_{x_t \in \mathcal{P}_d} \left(a_{t,i}^{\mathrm{probe}}\right)^2 =: \delta_i. \end{aligned} \tag{17}$$

Thus δ_i is the objective increase corresponding to teacher matching on the probe contexts. It is computed separately from the clipped calibration energy E_i^{cal} used to estimate component value.

A shared local response map. Let $\boldsymbol{\mathcal{J}} = (\mathcal{J}_i)_{i \in \mathcal{I}}$ and let G have rows $g_i^{\top} = \nabla_\theta \mathcal{J}_i(\theta_0)^{\top}$. For a fixed finite bank, differentiability at θ_0 gives

$$\boldsymbol{\mathcal{J}}(\theta_0 + \Delta\theta) - \boldsymbol{\mathcal{J}}(\theta_0) = G\Delta\theta + o(\|\Delta\theta\|_2). \tag{18}$$

All component responses are therefore described locally by a common linear map of the student's parameter update.

Using the fixed positive-definite metric M, define

$$u = M^{1/2}\Delta\theta, \qquad h_i = M^{-1/2} g_i, \qquad H = GM^{-1/2}. \tag{19}$$

The first-order responses and update cost become

$$G\Delta\theta = Hu, \qquad \frac{1}{2}\Delta\theta^{\top} M \Delta\theta = \frac{1}{2}\|u\|_2^2. \tag{20}$$

In particular, a step $u = \alpha h_j$, with $\alpha > 0$, produces first-order progress

$$h_i^{\top} u = \alpha h_i^{\top} h_j = \alpha g_i^{\top} M^{-1} g_j \tag{21}$$

on component i. The sign of this inner product determines whether the step advances, leaves unchanged, or opposes that component under the local response model. Update sharing is thus expressed through the geometry of the response gradients.

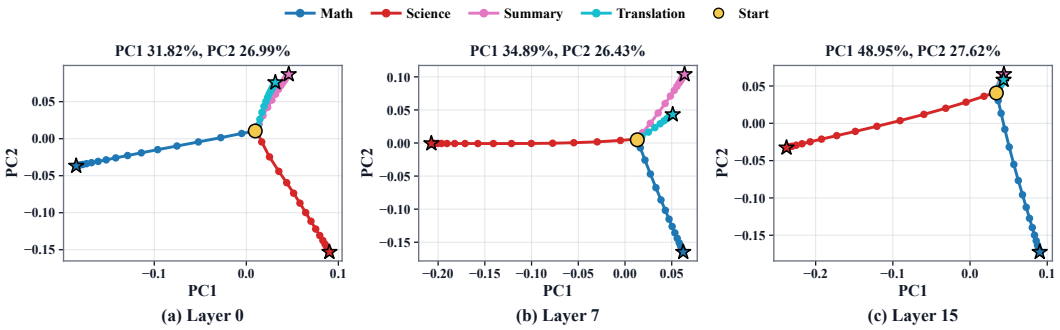


Figure 7: Domain-specific adaptation in common PCA coordinates within layers 0, 7, and 15. Dots show checkpoints, stars show final checkpoints, and the yellow marker denotes initialization. The first two components explain 58.81%, 61.32%, and 76.57% of checkpoint variation, respectively. Each panel describes the displayed layer's checkpoint geometry.

Adaptation trajectories. Figure 7 shows layerwise checkpoint trajectories from a common initialization. Translation and summarization follow relatively close projected paths; mathematics and science follow distinct paths.

A.2 Sufficiency of the Response Subspace (Q2)

Q1 expresses every selected first-order response through H. An update component in $\ker(H)$ therefore incurs norm cost without changing those responses. Removing this component yields an exact reduction of the local minimum-update problem.

Let $r = \operatorname{rank}(H)$ and let $Q \in \mathbb{R}^{p\times r}$ have orthonormal columns spanning $\operatorname{range}(H^\top)$. Define $A = HQ$ and the reduced update map

$$\Delta\theta(b) = M^{-1/2}Qb, \qquad b \in \mathbb{R}^r. \tag{22}$$

Proposition 1 (Exact reduction of local realization cost)**.** *For any selected set S, the cost in Equation 1 satisfies*

$$C(S) = \min_{b\in\mathbb{R}^r} \tfrac{1}{2}\|b\|_2^2 \quad \text{s.t.} \quad A_S b \geq \delta_S, \tag{23}$$

with identical feasibility in the original and reduced problems. For feasible S, its minimum-norm update belongs to $\operatorname{range}(H_S^\top)$.

Proof. Every $u \in \mathbb{R}^p$ decomposes uniquely as $u = Qb + u_\perp$, with $Q^\top u_\perp = 0$. Since every row of H lies in the span of Q,

$$Hu = HQb = Ab, \qquad \|u\|_2^2 = \|b\|_2^2 + \|u_\perp\|_2^2.$$

Consequently, any feasible full-space update gives a feasible reduced update of no greater cost. Conversely, each feasible b lifts to $u = Qb$ with the same responses and cost. Taking infima proves the equality and feasibility statement. For a nonempty feasible set, the closed feasible region and coercive objective ensure that the minimum is attained. Repeating the decomposition with $\operatorname{range}(H_S^\top)$ shows that any orthogonal component must vanish at the optimum. For $S = \varnothing$, the optimum is zero. □

The sufficient dimension is at most $\min(|S|, p)$ for one selected set and $\min(|\mathcal{I}|, p)$ for the common bank. Hence the local problem admits a lower-dimensional exact representation when its response rank is smaller than the parameter dimension. This result also preserves marginal costs for extensions of feasible sets. It uses a basis of the response span; JFMD instead uses a shared projection that is inexpensive to construct. Q3 examines the accuracy required of that projection.

A.3 Projection and Cost Preservation (Q3)

Randomized matrix approximation provides a general framework for representing high-dimensional structure in compact subspaces Halko et al. (2011). A practical sketch approximates the response geometry. For regularized response matching, its effect can be expressed through the error in the response Gram matrix. Define

$$K = HH^\top, \qquad \widetilde{K} = HP^\top PH^\top, \tag{24}$$

and the unprojected regularized cost

$$C_\eta(S) = \min_u \left\{ \tfrac{1}{2}\|u\|_2^2 + \tfrac{1}{2\eta}\|H_S u - \delta_S\|_2^2 \right\} = \tfrac{1}{2}\delta_S^\top (K_S + \eta I)^{-1}\delta_S. \tag{25}$$

The projected cost $\widehat{C}(S)$ has the same expression with $\widetilde{K}_S$ in place of K_S.

Proposition 2 (Cost error under Gram perturbation)**.** *For a fixed candidate bank and $\eta > 0$, suppose $\|\widetilde{K} - K\|_2 \le \xi$. For every S, define $b_S = \xi\|\delta_S\|_2^2/(2\eta^2)$, with $b_\varnothing = 0$. Then*

$$|\widehat{C}(S) - C_\eta(S)| \le b_S. \tag{26}$$

Proof. A principal submatrix satisfies $\|\widetilde{K}_S - K_S\|_2 \le \xi$. Both Gram matrices are positive semidefinite, so the inverse norms of $K_S + \eta I$ and $\widetilde{K}_S + \eta I$ are at most $1/\eta$. The resolvent identity gives

$$(\widetilde{K}_S + \eta I)^{-1} - (K_S + \eta I)^{-1} = (\widetilde{K}_S + \eta I)^{-1}(K_S - \widetilde{K}_S)(K_S + \eta I)^{-1}.$$

Its norm is at most ξ/η^2. Multiplying on both sides by δ_S and including the factor $1/2$ proves the result. □

Write $\Delta_\eta(i \mid S) = C_\eta(S \cup \{i\}) - C_\eta(S)$. The proposition immediately gives

$$|\widehat{\Delta C}_{i|S} - \Delta_\eta(i \mid S)| \le b_{S\cup\{i\}} + b_S. \tag{27}$$

Candidate comparisons admit a sharper condition because the base-set cost cancels.

Corollary 1 (Preservation of separated cost orderings)**.** *For $i, j \notin S$, if*

$$\Delta_\eta(j \mid S) - \Delta_\eta(i \mid S) > b_{S\cup\{i\}} + b_{S\cup\{j\}}, \tag{28}$$

then $\widehat{\Delta C}_{i|S} < \widehat{\Delta C}_{j|S}$.

Proof. The difference between two marginal costs equals the difference between their extended-set costs. Applying the proposition to these two sets bounds the error in their difference by $b_{S\cup\{i\}} + b_{S\cup\{j\}}$. The assumed strict gap therefore preserves its sign. □

Thus, projection accuracy matters relative to the cost separation of the candidates being compared. The bound holds simultaneously for subsets of the fixed bank, including adaptively selected ones. Here ξ denotes the spectral norm of Gram distortion. The conclusion concerns regularized cost ordering; value-per-cost scores and the absolute budget additionally determine selection.

Projection and ranking diagnostics. Table 4 evaluates retention under new projections and new examples. Increasing m from 64 to 1024 raises all-domain retention under a new projection from 0.426 to 0.908; retention on new examples rises from 0.364 to 0.643.

Table 5 compares surrogate and inequality-constrained QP marginal-cost rankings in the same projected space on Qwen3-1.7B banks. Across set sizes 1, 2, 4, and 8, five-seed mean Spearman correlations range from 0.9955 to 0.9998 for greedy prefixes and from 0.9962 to 0.9997 for random sets.

Table 4: Retention under new random projections and new examples across projection dimensions. Single, Pair, and All denote single-domain, domain-pair, and all-domain configurations. Higher values indicate greater retention.

	New projection			New examples		
m	Single	Pair	All	Single	Pair	All
64	0.519	0.445	0.426	0.427	0.371	0.364
128	0.706	0.631	0.584	0.587	0.483	0.456
256	0.792	0.739	0.710	0.683	0.607	0.564
512	0.902	0.869	0.855	0.728	0.659	0.642
1024	0.934	0.918	0.908	0.737	0.676	0.643

Table 5: Marginal-cost ranking validation on Qwen3-1.7B calibration banks, with 32 candidate extensions per set. Entries report the mean $\pm$ sample standard deviation across five seeds, computed from within-seed averages. Each seed contributes one greedy prefix per size. For random sets, each seed contributes 62–63 sets at size 1 and 100 sets at each larger size. QP percentile gives the rank of the surrogate's lowest-cost candidate under the projected QP reference on a 0–100 scale (0 is best).

Set source	Set size	Spearman ρ	QP percentile $\downarrow$
Greedy prefixes	1	0.9998 ± 0.0004	0.00 ± 0.00
	2	0.9991 ± 0.0008	0.00 ± 0.00
	4	0.9955 ± 0.0042	0.13 ± 0.10
	8	0.9966 ± 0.0026	0.00 ± 0.00
Random cross-domain	1	0.9997 ± 0.0001	0.05 ± 0.09
	2	0.9996 ± 0.0002	0.16 ± 0.17
	4	0.9996 ± 0.0002	0.27 ± 0.00
	8	0.9962 ± 0.0017	1.10 ± 0.83

A.4 Residual Reconstruction and Mode Energy (Q4)

The preceding analysis concerns the update cost of supervision components. We now examine the residual representation used to estimate their value.

For domain d, let $r_{d,t}(\theta_0)$ be the initial log-probability residual on the anchor vocabulary. Mode construction uses its elementwise-clipped form

$$\bar{r}_{d,t} = \mathrm{clip}\left(r_{d,t}(\theta_0), -c, c\right), \qquad c = 20. \tag{29}$$

Stack these vectors as rows of $\bar{R}_d \in \mathbb{R}^{N_d \times |\mathcal{V}_d|}$. Its compact SVD is

$$\bar{R}_d = \sum_{k=1}^{r_d} \sigma_{d,k} u_{d,k} v_{d,k}^\top, \qquad \sigma_{d,1} \geq \cdots \geq \sigma_{d,r_d} > 0, \tag{30}$$

where $r_d = \mathrm{rank}(\bar{R}_d)$. The candidate bank contains the first $K_d \leq r_d$ right singular vectors. Since $\bar{R}_d v_{d,k} = \sigma_{d,k} u_{d,k}$, the calibration energy is

$$E_{d,k}^{\mathrm{cal}} = \frac{\|\bar{R}_d v_{d,k}\|_2^2}{N_d} = \frac{\sigma_{d,k}^2}{N_d}. \tag{31}$$

It measures the mean squared clipped residual captured along the mode direction.

Proposition 3 (Residual energy decomposition)**.** *For* $S_d \subseteq \{1, \ldots, K_d\}$, *let* $\Pi_{S_d} = \sum_{k \in S_d} v_{d,k} v_{d,k}^\top$. *Then*

$$\frac{\|\bar{R}_d \Pi_{S_d}\|_F^2}{N_d} = \sum_{k \in S_d} E_{d,k}^{\mathrm{cal}}, \tag{32}$$

$$\frac{\|\bar{R}_d - \bar{R}_d \Pi_{S_d}\|_F^2}{N_d} = \sum_{\substack{k \leq K_d \\ k \notin S_d}} E_{d,k}^{\mathrm{cal}} + \frac{1}{N_d} \sum_{k=K_d+1}^{r_d} \sigma_{d,k}^2. \tag{33}$$

Moreover, $\bar{R}_d\Pi_{S_d}$ minimizes the Frobenius reconstruction error among matrices whose rows lie in the selected span.

Proof. The rank-one matrices $u_{d,k}v_{d,k}^\top$ are orthonormal under the Frobenius inner product. Projection retains exactly the terms indexed by S_d, so their squared norms add. The omitted terms consist of unselected bank modes and singular directions beyond K_d, giving the two identities.

For any matrix X whose rows lie in the selected span, orthogonality gives

$$\|\bar{R}_d - X\|_F^2 = \|\bar{R}_d(I - \Pi_{S_d})\|_F^2 + \|\bar{R}_d\Pi_{S_d} - X\|_F^2.$$

The second term is minimized at $X = \bar{R}_d\Pi_{S_d}$. □

For the complete candidate bank, the error is

$$\frac{\|\bar{R}_d - \bar{R}_d\Pi_{\{1,\dots,K_d\}}\|_F^2}{N_d} = \frac{1}{N_d}\sum_{k=K_d+1}^{r_d}\sigma_{d,k}^2. \tag{34}$$

The decomposition separates bank-truncation error from the additional error introduced by mode selection. For a nonzero-energy bank,

$$e_{d,k} = \frac{E_{d,k}^{\mathrm{cal}}}{\sum_{j=1}^{K_d} E_{d,j}^{\mathrm{cal}}}$$

gives unit total energy share before reliability adjustment.

These identities concern the clipped calibration matrix used to construct the bank. The response-matching target δ_i is instead computed from unclipped residuals on the probe contexts, as specified in Appendix C.3. Calibration energy and target progress therefore have distinct roles in the selection framework. The decomposition gives calibration energy a reconstruction-based interpretation. Appendix D.1 compares energy-based selection with the complete selector.

A.5 Cross-Sample Residual Stability (Q5)

Q4 characterizes reconstruction on the calibration sample. We now relate that sample representation to the residual distribution under the same preprocessing. Fix a domain, the initial student, and the anchor vocabulary. Let $r \in \mathbb{R}^D$ denote the corresponding log-probability residual, and define $\bar{r} = \mathrm{clip}(r, -c, c)$ with the same $c = 20$ used in calibration. The population and empirical uncentered second moments are

$$\Sigma = \mathbb{E}[\bar{r}\bar{r}^\top], \qquad \widehat{\Sigma} = \frac{1}{N}\sum_{t=1}^{N}\bar{r}_t\bar{r}_t^\top, \qquad \zeta = \|\widehat{\Sigma} - \Sigma\|_2. \tag{35}$$

These uncentered second moments match the residual decomposition. Coordinatewise clipping ensures $\|\bar{r}\|_2^2 \le Dc^2$, so the population second moment is finite. The following results express stability in terms of the realized second-moment discrepancy ζ.

Proposition 4 (Uniform stability of captured energy)**.** *For every unit vector v and rank-K orthogonal projector Π,*

$$\left|v^\top\widehat{\Sigma}v - v^\top\Sigma v\right| \le \zeta, \tag{36}$$

$$\left|\mathrm{tr}(\Pi\widehat{\Sigma}) - \mathrm{tr}(\Pi\Sigma)\right| \le K\zeta. \tag{37}$$

The bounds also hold for directions and projectors selected from the calibration sample.

Proof. For any unit vector v,

$$\left|v^\top(\widehat{\Sigma} - \Sigma)v\right| \le \|\widehat{\Sigma} - \Sigma\|_2 = \zeta.$$

Write $\Pi = \sum_{k=1}^{K} v_k v_k^\top$ with orthonormal v_k. Summing the corresponding quadratic-form bounds gives the trace inequality. Both bounds hold simultaneously over all such vectors and projectors, including data-dependent choices. □

Captured energy can remain stable even when individual directions are not uniquely identified. To control the top-K subspace itself, we also consider the spectral separation at the bank boundary. Let $\Pi_K^\star$ and $\widehat{\Pi}_K$ be top-K spectral projectors of Σ and $\widehat{\Sigma}$, respectively. Define the population reconstruction error as

$$\mathcal{R}(\Pi) = \mathbb{E}\|\bar{r} - \Pi\bar{r}\|_2^2.$$

Proposition 5 (Reconstruction and subspace persistence)**.** *For $1 \le K \le D$,*

$$0 \le \mathcal{R}(\widehat{\Pi}_K) - \mathcal{R}(\Pi_K^\star) \le 2K\zeta. \tag{38}$$

If $K < D$ and $\gamma = \lambda_K(\Sigma) - \lambda_{K+1}(\Sigma) > 0$, then

$$\|\widehat{\Pi}_K - \Pi_K^\star\|_F^2 \le \frac{4K\zeta}{\gamma}. \tag{39}$$

Proof. For an orthogonal projector,

$$\mathcal{R}(\Pi) = \operatorname{tr}(\Sigma) - \operatorname{tr}(\Pi\Sigma).$$

Population optimality of $\Pi_K^\star$ gives nonnegativity of the reconstruction difference. Empirical optimality of $\widehat{\Pi}_K$ gives

$$\operatorname{tr}\big((\Pi_K^\star - \widehat{\Pi}_K)\widehat{\Sigma}\big) \le 0.$$

Consequently,

$$\begin{aligned}\operatorname{tr}\big((\Pi_K^\star - \widehat{\Pi}_K)\Sigma\big) &\le \operatorname{tr}\big(\Pi_K^\star(\Sigma - \widehat{\Sigma})\big) + \operatorname{tr}\big(\widehat{\Pi}_K(\widehat{\Sigma} - \Sigma)\big) \\ &\le 2K\zeta,\end{aligned}$$

where the final inequality follows from Equation 37.

For the subspace bound, let v_j be an orthonormal eigenbasis of Σ, with eigenvalues in nonincreasing order, and write $a_j = v_j^\top \widehat{\Pi}_K v_j$. Then $0 \le a_j \le 1$ and $\sum_j a_j = K$. Hence

$$\begin{aligned}\operatorname{tr}\big((\Pi_K^\star - \widehat{\Pi}_K)\Sigma\big) &= \sum_{j \le K} \lambda_j(1 - a_j) - \sum_{j > K} \lambda_j a_j \\ &\ge \gamma \sum_{j \le K} (1 - a_j) \\ &= \frac{\gamma}{2}\|\widehat{\Pi}_K - \Pi_K^\star\|_F^2.\end{aligned}$$

Combining this inequality with the reconstruction bound proves the result. □

The first result controls captured energy through second-moment accuracy. The second additionally requires spectral separation to identify the top-K bank subspace. Individual singular vectors can rotate within a tied eigenspace, motivating evaluation of a candidate through its representation in split subspaces rather than a one-to-one match between singular vectors. JFMD combines this directional retention with split-energy consistency, as defined in Appendix C.2. The bounds describe the clipped residual representation; the empirical analysis below examines how the implemented reliability score relates to the reported held-out stability measure.

Independent-sample evidence. Table 6 reports the association between calibration reliability and held-out stability across 192 modes. The Spearman correlation is 0.680 and remains positive at 0.288 after controlling for energy. Across 221 energy-matched pairs, the higher-reliability mode has a mean stability advantage of 0.052. The partial correlation and energy-matched comparisons associate reliability with held-out stability beyond calibration energy.

At $m = 512$ and 1024, all-domain retention on new examples is 0.642 and 0.643, respectively, while new-projection retention increases from 0.855 to 0.908 (Table 4). The two diagnostics separate projection reproducibility from cross-sample persistence.

Table 6: Association between calibration reliability and held-out mode stability. The analyses use 192 modes and 221 energy-matched pairs. The regression coefficient uses the fitted model's scale. The matched difference subtracts the lower-reliability mode's stability from that of the higher-reliability mode.

Analysis	Result
Spearman correlation	0.680
Partial rank correlation controlling for energy	0.288
Reliability coefficient controlling for $\log E_i^{\mathrm{cal}}$ and seed–domain fixed effects	0.833
Mean stability difference across 221 energy-matched pairs	+0.052

B Joint Cost and Set-Conditional Selection

Building on the response representations in Appendix A, we examine how supervision components interact through shared student updates and how these interactions guide selection. The analysis follows four questions:

Q6. How does progress supplied by the selected components affect the cost of adding another?

Q7. Why can candidate cost rankings change as the selected set grows?

Q8. How can joint costs and their marginal increments be computed efficiently?

Q9. How are component values and marginal costs combined into budget-feasible selection?

B.1 Minimum-Update Cost and Update Reuse (Q6)

The response representation in Appendix A.1 allows us to express supervision requirements in a common update space. We now examine how the update already required by a selected set affects the cost of adding another component.

Progress supplied by the selected set. Using the whitened coordinates $u = M^{1/2}\Delta\theta$ and $h_i = M^{-1/2}g_i$, the local realization cost is

$$C(S) = \min_u \frac{1}{2}\|u\|_2^2 \quad \text{s.t.} \quad h_j^\top u \geq \delta_j, \qquad j \in S. \tag{40}$$

For feasible S, the objective has a unique minimizer $u_S^\star$. Before adding candidate $i \notin S$, this update already supplies progress $h_i^\top u_S^\star$. The remaining requirement is therefore

$$d_{i|S} = [\delta_i - h_i^\top u_S^\star]_+, \qquad [a]_+ = \max(a, 0). \tag{41}$$

If this quantity is zero, the existing update also satisfies the candidate. Otherwise, additional progress must be obtained while maintaining feasibility for the selected components.

Let

$$A = \{j \in S : h_j^\top u_S^\star = \delta_j\}$$

be the active constraint set, and let Π_A denote the orthogonal projector onto $\mathrm{span}\{h_j : j \in A\}$, with $\Pi_A = 0$ when $A = \varnothing$.

Proposition 6 (Marginal cost under update reuse)**.** *For feasible S and $i \notin S$,*

$$\Delta C_{i|S} = 0 \quad \Longleftrightarrow \quad d_{i|S} = 0.$$

If $d_{i|S} > 0$ and the expanded problem has an optimum whose active set is exactly $A \cup \{i\}$, with linearly independent active constraint vectors, then

$$\Delta C_{i|S} = \frac{d_{i|S}^2}{2\|(I - \Pi_A)h_i\|_2^2}. \tag{42}$$

Proof. For feasible S, the feasible region is closed and convex, and the objective is coercive and strictly convex. Hence $u_S^\star$ exists and is unique.

If $d_{i|S} = 0$, then $u_S^\star$ is also feasible for $S \cup \{i\}$. Since adding a constraint cannot reduce the minimum cost,

$$C(S) \le C(S \cup \{i\}) \le \frac{1}{2}\|u_S^\star\|_2^2 = C(S).$$

Thus $\Delta C_{i|S} = 0$. Conversely, if the marginal cost is zero, the expanded problem has the same finite optimum value. Its minimizer is feasible and optimal for the original problem, so uniqueness implies $u_{S\cup\{i\}}^\star = u_S^\star$. The candidate constraint is therefore already satisfied, giving $d_{i|S} = 0$.

Now suppose $d_{i|S} > 0$ and the stated active-set conditions hold. Define

$$v = (I - \Pi_A)h_i, \qquad w = u_{S\cup\{i\}}^\star - u_S^\star.$$

Linear independence gives $v \neq 0$. The KKT stationarity conditions imply

$$u_S^\star \in \operatorname{span}\{h_j : j \in A\}, \qquad u_{S\cup\{i\}}^\star \in \operatorname{span}\big(\{h_j : j \in A\} \cup \{h_i\}\big).$$

Every constraint in A is active at both optima, so $h_j^\top w = 0$ for all $j \in A$. Consequently, w lies in the one-dimensional span of v.

Because the candidate constraint is active at the expanded optimum,

$$h_i^\top w = \delta_i - h_i^\top u_S^\star = d_{i|S}.$$

Using $h_i^\top v = \|v\|_2^2$, we obtain

$$w = \frac{d_{i|S}}{\|v\|_2^2} v.$$

Moreover, $w \perp u_S^\star$. Therefore,

$$\begin{aligned}\Delta C_{i|S} &= \frac{1}{2}\|u_S^\star + w\|_2^2 - \frac{1}{2}\|u_S^\star\|_2^2 \\ &= \frac{1}{2}\|w\|_2^2 = \frac{d_{i|S}^2}{2\|(I - \Pi_A)h_i\|_2^2}.\end{aligned}$$

□

The expression separates two aspects of adding a component: how much progress remains, and how strongly the candidate responds along directions that leave the existing active responses unchanged. Under the stated active-set conditions, greater supplied progress reduces the numerator, while a larger available response reduces the update needed to complete the target. When the active set changes, the marginal cost remains defined by the two optimization problems, with the new active geometry determining its value.

Controlled geometry and numerical evidence. The consequences of update reuse can be seen directly in two-component configurations. Consider $\|h_1\|_2 = \|h_2\|_2 = a > 0$ and equal targets $\delta_1 = \delta_2 = \delta > 0$. Each component alone has cost

$$C(\{i\}) = \frac{\delta^2}{2a^2}.$$

To compare joint and separate realization, define

$$\kappa = \frac{C(\{1,2\})}{C(\{1\}) + C(\{2\})}, \qquad \Gamma = \log \frac{\rho_{\{1,2\}}}{\sqrt{\rho_{\{1\}}^2 + \rho_{\{2\}}^2}} = \frac{1}{2}\log \kappa, \tag{43}$$

where $\rho_T = \sqrt{2C(T)}$ is the minimum update norm. Thus κ measures the quadratic-cost ratio, while Γ expresses the same comparison on a logarithmic update-norm scale.

For duplicate requirements, $h_2 = h_1$, the first component's optimum satisfies both constraints. Hence $C(\{1,2\}) = C(\{1\})$, giving $\kappa = 1/2$ and $\Gamma = -\frac{1}{2}\log 2$. The second requirement has zero marginal cost.

For orthogonal requirements, $h_1^\top h_2 = 0$, the optimum is $u^\star = \delta(h_1 + h_2)/a^2$. The squared norms add, giving $\kappa = 1$ and $\Gamma = 0$. The second component requires its full standalone cost.

Table 7: Controlled feasibility validation with 12 experiments per configuration. Finite numerical results use the update-norm surcharge Γ; the quadratic-cost ratio is κ. For sign-flipped targets, the numerical entry reports the number of feasible joint problems under the tested strict realization threshold.

Configuration	Analytical reference	Numerical result
Duplicate	$\kappa = 1/2, \Gamma \approx -0.3466$	$\Gamma \approx -0.3463$
Block-diagonal	$\kappa = 1, \Gamma = 0$	$\Gamma \approx 0.0003$
Sign-flipped	Jointly infeasible	Feasible: $0/12$

For opposing requirements, $h_2 = -h_1$, simultaneous positive progress would imply

$$h_1^\top u \geq \delta, \qquad -h_1^\top u \geq \delta.$$

Adding the inequalities gives $0 \geq 2\delta$, so the joint problem is infeasible although each component is individually feasible.

Figure 2(a) illustrates these three regimes. Table 7 reports the corresponding controlled numerical diagnostics from the study, with 12 experiments per configuration. Duplicate targets yield a surcharge of approximately -0.3463, compared with the analytical value -0.3466; block-diagonal targets yield approximately 0.0003, compared with zero. For sign-flipped targets, none of the 12 joint problems is feasible under the tested strict realization threshold.

The three configurations have identical standalone costs but different joint requirements. Their numerical behavior agrees with the predicted effects of reuse, independence, and incompatibility. This distinction motivates evaluating a candidate through the additional update required by the current set. The next subsection establishes how it can change the ordering of candidate components.

B.2 Set-Dependent Candidate Ordering (Q7)

The update-reuse result in Appendix B.1 implies that a component's standalone cost need not reflect the additional cost of learning it after other components have been selected. This distinction can change which candidate is less expensive to add.

Why candidate orderings can change. For a nonzero response vector h_i and target $\delta_i > 0$, the singleton problem has solution

$$u^\star_{\{i\}} = \frac{\delta_i}{\|h_i\|_2^2} h_i, \qquad C(\{i\}) = \frac{\delta_i^2}{2\|h_i\|_2^2}. \tag{44}$$

This cost depends only on the component's own response and target. Its marginal cost additionally accounts for progress supplied by the selected set.

Corollary 2 (Set-dependent candidate ordering)**.** *There exist a feasible set S and candidates $i, j \notin S$ such that*

$$C(\{i\}) < C(\{j\}), \qquad \Delta C_{i|S} > \Delta C_{j|S}. \tag{45}$$

Proof. Consider two-dimensional Euclidean update coordinates with unit progress targets. Let $S = \{s\}$, where

$$h_s = (1, 0)^\top, \qquad \delta_s = 1.$$

The optimal update is $u^\star_S = (1, 0)^\top$, with $C(S) = 1/2$.

Take distinct candidate components with

$$h_i = (0, 2)^\top, \qquad h_j = (1, 0)^\top, \qquad \delta_i = \delta_j = 1.$$

Their standalone costs satisfy

$$C(\{i\}) = \frac{1}{8} < \frac{1}{2} = C(\{j\}).$$

After conditioning on S, the current update already satisfies candidate j, giving $\Delta C_{j|S} = 0$. Candidate i requires the additional constraint $u_2 \geq 1/2$. The expanded optimum is $(1, 1/2)^\top$, so

$$\Delta C_{i|S} = \frac{1}{2}\left(1 + \frac{1}{4}\right) - \frac{1}{2} = \frac{1}{8}.$$

Hence the standalone ordering reverses. □

The construction reverses candidate ordering through update reuse: S already supplies the progress required by j. Figure 2(c) shows the corresponding set dependence in the calibrated bank: mean Kendall agreement between singleton and marginal-cost rankings falls from approximately 0.90 to 0.60 as set size increases from 2 to 8.

Realized update requirements. To examine whether the ranking difference corresponds to actual learning difficulty, we evaluate multiple candidate pairs for which singleton cost favors i and conditional cost given the selected set S favors j. For each pair, both candidate branches start from the same student adapted to S. We compare their learning progress at matched parameter-update cost, using continued training on S alone as the reference for changes in the previously learned modes. We aggregate the comparisons across candidate pairs and report variation across seeds.

For mode k, let $e_k(t)$ be the mean squared projection of the teacher–student log-probability residual onto its mode direction on held-out contexts at checkpoint t. We measure candidate progress and realized update cost as

$$P_k(t) = 1 - \frac{e_k(t)}{e_k(0)}, \qquad u(t) = \left(\sum_{p \in \text{trainable}} \|\theta_p(t) - \theta_p(0)\|_2^2\right)^{1/2}, \tag{46}$$

where $\theta(0)$ is the common branch starting point after adaptation to S. Thus $u(t)$ measures Euclidean displacement of the trainable LoRA parameters from that point. Below, $P_k(u)$ denotes progress along the observed training trajectory indexed by this cost. We measure degradation on the selected set by

$$H_k(u) = \max_{s \in S} \left[P_s^S(u) - P_s^{S+k}(u)\right], \tag{47}$$

where the two progress values are evaluated at the same update cost. Positive values indicate that at least one selected mode performs worse than in the only-S branch. The update required to reach a candidate progress target while satisfying the preservation condition is

$$R_k(\tau, \epsilon) = \inf\left\{u : P_k(u) \geq \tau,\ H_k(u) \leq \epsilon\right\}, \tag{48}$$

where ϵ is a fixed preservation tolerance. For brevity, the main text suppresses the dependence of R_k on τ and ϵ. The quantity R_k measures update requirements along the observed training path under the preservation tolerance ϵ.

At progress targets of 0.4, 0.5, and 0.6, the aggregate update requirements across the evaluated candidate pairs are 10.9%, 42.9%, and 34.4% lower for candidates preferred by conditional cost than for those preferred by singleton cost, respectively (Table 1). This aggregate ordering is consistent across the five seeds.

B.3 Tractable Joint Cost and Marginal Increments (Q8)

The preceding analysis motivates evaluating each candidate relative to the selected set. For tractable selection, JFMD uses regularized response matching in the shared projected space. We derive the surrogate's closed-form set cost, exact marginal increments, and incremental updates for efficient selection.

Regularized joint cost. Let $\widetilde{g}_i = PM^{-1/2}g_i$ be the projected response of component i. For a selected set S, let $\widetilde{G}_S$ stack the row vectors $\widetilde{g}_i^\top$ and let δ_S collect their progress targets. Define

$$F_S(z) = \frac{1}{2}\|z\|_2^2 + \frac{1}{2\eta}\|\widetilde{G}_S z - \delta_S\|_2^2, \qquad \widehat{C}(S) = \min_z F_S(z), \quad \eta > 0. \tag{49}$$

The objective balances update size with the remaining response-matching error. Its Hessian is

$$A_S = I + \frac{1}{\eta}\widetilde{G}_S^\top \widetilde{G}_S \succ 0, \tag{50}$$

so the minimizer $z_S^\star$ is unique.

Proposition 7 (Closed-form response-matching cost)**.** *For fixed projected responses, targets, and $\eta > 0$,*

$$z_S^\star = \widetilde{G}_S^\top \left(\widetilde{G}_S \widetilde{G}_S^\top + \eta I\right)^{-1} \delta_S, \tag{51}$$

$$\widehat{C}(S) = \frac{1}{2}\delta_S^\top \left(\widetilde{G}_S \widetilde{G}_S^\top + \eta I\right)^{-1} \delta_S. \tag{52}$$

For $S = \varnothing$, $z_S^\star = 0$ and $\widehat{C}(S) = 0$.

Proof. Write $G = \widetilde{G}_S$, $\delta = \delta_S$, and $K = GG^\top + \eta I$. The first-order condition is

$$(G^\top G + \eta I)z = G^\top \delta.$$

Since

$$(G^\top G + \eta I)G^\top = G^\top (GG^\top + \eta I),$$

the unique solution is $z_S^\star = G^\top K^{-1}\delta$.

Let $\alpha = K^{-1}\delta$. The response residual at this solution satisfies

$$Gz_S^\star - \delta = GG^\top \alpha - \delta = -\eta\alpha.$$

Substitution into the objective gives

$$\begin{aligned} F_S(z_S^\star) &= \frac{1}{2}\alpha^\top GG^\top \alpha + \frac{\eta}{2}\alpha^\top \alpha \\ &= \frac{1}{2}\alpha^\top K\alpha = \frac{1}{2}\delta^\top K^{-1}\delta. \end{aligned}$$

The empty-set result follows directly from $F_\varnothing(z) = \frac{1}{2}\|z\|_2^2$. □

The Gram matrix records both response strength and alignment between components. Consequently, the cost depends on the responses jointly rather than on a sum of independent component costs.

Exact marginal increments. For candidate $i \notin S$, define its signed response mismatch at the current optimum as

$$r_{i|S} = \delta_i - \widetilde{g}_i^\top z_S^\star, \qquad v_{i|S} = \widetilde{g}_i^\top A_S^{-1}\widetilde{g}_i. \tag{53}$$

The first quantity measures how far the current response is from the candidate's target. The second measures the candidate's response relative to the curvature of the current objective.

Proposition 8 (Exact marginal response-matching cost)**.** *For every S and $i \notin S$,*

$$\widehat{\Delta C}_{i|S} = \frac{r_{i|S}^2}{2(\eta + v_{i|S})}. \tag{54}$$

The expanded optimum is

$$z_{S\cup\{i\}}^\star = z_S^\star + \frac{r_{i|S}}{\eta + v_{i|S}} A_S^{-1}\widetilde{g}_i. \tag{55}$$

In particular, $\widehat{C}$ is monotone under set inclusion, and $\widehat{\Delta C}_{i|S} = 0$ exactly when $r_{i|S} = 0$.

Proof. Because F_S is quadratic with Hessian A_S, expansion around its minimizer gives

$$F_S(z_S^\star + w) = \widehat{C}(S) + \frac{1}{2}w^\top A_S w.$$

Adding candidate i therefore yields

$$\widehat{\Delta C}_{i|S} = \min_{w} \left\{ \frac{1}{2} w^\top A_S w + \frac{1}{2\eta} (\widetilde{g}_i^\top w - r_{i|S})^2 \right\}.$$

Its stationarity equation is

$$A_S w + \frac{1}{\eta} \widetilde{g}_i (\widetilde{g}_i^\top w - r_{i|S}) = 0.$$

The unique solution is

$$w^\star = \frac{r_{i|S}}{\eta + v_{i|S}} A_S^{-1} \widetilde{g}_i.$$

At this solution,

$$(w^\star)^\top A_S w^\star = \frac{r_{i|S}^2 v_{i|S}}{(\eta + v_{i|S})^2},$$

and

$$\widetilde{g}_i^\top w^\star - r_{i|S} = -\frac{\eta r_{i|S}}{\eta + v_{i|S}}.$$

Substituting both expressions into the objective gives Equation 54. Since $A_S \succ 0$, we have $v_{i|S} \geq 0$. The denominator is therefore positive, proving nonnegativity and the zero-cost characterization. Monotonicity under set inclusion follows by successively adding components. □

The surrogate increment depends on the selected set through the current response mismatch and the curvature of the response-matching objective. Regularized matching penalizes both undershooting and overshooting, so its marginal cost is zero only when the current response exactly matches the candidate target. By contrast, for a feasible set, the hard-constraint increment in Appendix B.1 is zero whenever its minimum-cost update meets or exceeds that target. Thus $\Delta\widehat{C}_{i|S}$ is the exact increase in the regularized response-matching objective used for selection.

Incremental evaluation. The marginal expression avoids solving a new optimization problem for every candidate. After accepting component i,

$$A_{S\cup\{i\}} = A_S + \frac{1}{\eta} \widetilde{g}_i \widetilde{g}_i^\top.$$

The Sherman–Morrison identity gives

$$A_{S\cup\{i\}}^{-1} = A_S^{-1} - \frac{A_S^{-1} \widetilde{g}_i \widetilde{g}_i^\top A_S^{-1}}{\eta + v_{i|S}}. \tag{56}$$

Together with Equation 55, this provides an exact incremental evaluation scheme for the surrogate in exact arithmetic. The initial state is $A_\varnothing^{-1} = I$, $z_\varnothing^\star = 0$, and $\widehat{C}(\varnothing) = 0$.

For comparison, a component's singleton cost is

$$\widehat{C}(\{i\}) = \frac{\delta_i^2}{2(\eta + \|\widetilde{g}_i\|_2^2)}. \tag{57}$$

Conditioning on S replaces the initial target mismatch and Euclidean response norm with $r_{i|S}$ and $v_{i|S}$. These are precisely the quantities that evolve as other supervision components are retained.

These exact marginal increments supply the cost term in the selector below. Table 5 evaluates their ranking agreement with the projected inequality-constrained QP.

B.4 Supervision Selection with Domain Coverage (Q9)

The preceding derivations provide the cost of adding a component to the selected set. The implementation uses these increments in two stages: domain-coverage initialization followed by global selection. Both stages evaluate candidates relative to the current set, while the budget check applies to the global stage.

Eligible candidates and coverage. The implemented eligibility test is

$$\mathcal{I}_{\text{elig}} = \{i : \rho_i \geq r_{\min}\}. \tag{58}$$

In the unnormalized cost units of Equation 10, the nominal budget is

$$B = \max(b, 0) \sum_{i \in \mathcal{I}_{\text{elig}}} \widehat{C}(\{i\}). \tag{59}$$

The implementation divides both set costs and the budget by the mean positive raw singleton cost over the full candidate bank, before reliability filtering. This common scaling leaves the budget comparisons in the unnormalized notation below unchanged. The implementation applies numerical floors when evaluating mode values and cost ratios, rather than a separate positive-energy eligibility test. Let $\text{Score}_{\text{impl}}(i \mid S)$ denote the value-per-marginal-cost score with the configured powers and numerical floors. Its unit-power form corresponds to Equation 11.

For each domain d, let m_d be the requested coverage count and c_d its selection cap. An inactive cap can be represented by the size of that domain's candidate bank. Starting from $S = \varnothing$, domains are processed in a fixed order. Within domain d, the procedure repeatedly selects the highest-scoring remaining eligible mode, recomputing scores after each addition, until

$$\min\{m_d,\ |\mathcal{I}_{\text{elig}} \cap \mathcal{I}_d|,\ c_d\}$$

modes have been retained.

Coverage additions are charged through the same set cost $\widehat{C}(S)$, but are not rejected when they exceed the nominal budget B. The reliability threshold is not relaxed when a domain has insufficient eligible modes. Denote the resulting set by S_{cov}.

Global selection. After coverage initialization, define

$$B_{\text{eff}} = \max\{B, \widehat{C}(S_{\text{cov}})\}. \tag{60}$$

Let $\mathcal{R}(S)$ contain the remaining eligible candidates whose addition respects the configured domain caps. The budget-feasible extensions are

$$\mathcal{F}(S) = \left\{ i \in \mathcal{R}(S) : \widehat{C}(S) + \widehat{\Delta C}_{i|S} \leq B_{\text{eff}} \right\}. \tag{61}$$

When this set is nonempty, the next component is

$$i^\star \in \arg\max_{i \in \mathcal{F}(S)} \text{Score}_{\text{impl}}(i \mid S). \tag{62}$$

The procedure adds $i^\star$, updates the joint cost and response-matching state, and recomputes the remaining candidate scores. It terminates when $\mathcal{F}(S) = \varnothing$.

In exact arithmetic, this fixed effective budget is equivalent to the implementation's repeated check against $\max\{B, \widehat{C}(S)\}$. If coverage exceeds B, monotonicity permits only zero-cost additions thereafter.

Selection procedure. Selection starts from $S = \varnothing$ and processes domains in a fixed order. Within each domain, eligible modes are added according to their current scores until the coverage count is met, no eligible candidates remain, or the selection cap is reached. The resulting set defines S_{cov}, and the effective budget is fixed at $B_{\text{eff}} = \max\{B, \widehat{C}(S_{\text{cov}})\}$. Selection then repeatedly evaluates the marginal costs of the remaining admissible candidates and constructs $\mathcal{F}(S)$. If $\mathcal{F}(S)$ is empty, the procedure returns S. Otherwise, it adds the highest-scoring candidate, updates the response-matching state, and reevaluates the remaining candidates.

The exact marginal expression in Appendix B.3 supports these repeated evaluations. Its incremental identities provide one algebraically equivalent way to maintain the response-matching state.

Budget control and termination.

Algorithm 1 Jointly Feasible Mode Distillation

Require: Domain teachers, initialized student, calibration and probe data, training data, nominal cost budget B
Ensure: Trained student
Construct supervision modes
for each domain **do**
  Compute teacher–student residuals on calibration data
  Extract residual modes using truncated SVD
  Estimate mode values from energy and cross-sample reliability
  Estimate mode response gradients and progress targets on probe data
  Project response gradients into a shared low-dimensional space
end for
Select jointly feasible modes
Remove modes below the reliability threshold
Initialize S with eligible modes for domain coverage
Fix $B_{\text{eff}} \leftarrow \max\{B, \widehat{C}(S)\}$
while an unselected eligible mode fits within B_{eff} **do**
  Recompute marginal surrogate costs given the current set S
  Add the budget-feasible mode with the highest value-to-marginal-cost score
end while
Freeze the selected mode bases
Distill selected corrections
for each training minibatch **do**
  Compute current teacher–student residuals
  Project residuals onto the selected modes
  Add the filtered corrections to student logits to form stop-gradient targets
  Update the student using supervised and distillation losses
end for

Proposition 9 (Budget control after coverage)**.** *For a finite candidate bank and exact cost evaluation, the global selection stage maintains*

$$\widehat{C}(S) \le B_{\text{eff}}$$

and terminates after finitely many additions. If $\widehat{C}(S_{\text{cov}}) \le B$, the returned set also satisfies the nominal budget $\widehat{C}(S) \le B$.

Proof. The coverage set satisfies $\widehat{C}(S_{\text{cov}}) \le B_{\text{eff}}$ by definition. Each subsequent addition is accepted only when its extended-set cost is at most B_{eff}, so the bound follows by induction. Each addition selects a previously unselected candidate from a finite bank, proving termination.

If coverage satisfies the nominal budget, then $B_{\text{eff}} = B$. Otherwise, $B_{\text{eff}} = \widehat{C}(S_{\text{cov}})$. Since the set cost is monotone, accepted additions in this case have zero marginal cost. □

The returned set defines the fixed correction subspace used in training; its residual coefficients are recomputed as the student changes (Section 4).

Algorithm 1 summarizes mode construction, joint-cost selection, and mode-filtered distillation.

C Implementation and Experimental Protocols

This section specifies calibration, response estimation, target construction, training, and evaluation.

C.1 Teacher–Student Configurations and Data

We train one shared student with supervision from four domain-specialized teachers. The principal configurations are Qwen3-8B→Qwen3-1.7B, Llama-3.1-8B→Llama-3.2-1B, and Gemma-7B→Gemma-2B. Each domain teacher is obtained by adapting its backbone to the corresponding domain. Within each configuration, comparisons use the same student initialization, domain teachers, source training data, and train/validation split.

The four domains are mathematics, science, translation, and summarization. Main experiments assign each example to its teacher using the known domain label. The shared student is evaluated on GSM8K, MATH, SciQ, OpenBookQA, FLORES, IWSLT, XSum, and SAMSum, using the metrics in Section 5.1.

Teacher-forced records contain a prompt, its domain label, and a reference response. CE and KD losses use valid response positions; prompt and padding positions are excluded. Validation data are reserved for checkpoint selection.

Residual construction uses token coordinates within each teacher–student pair.

C.2 Calibration and Mode Construction

Calibration residuals. The mode bank is constructed once at θ_0. The reported calibration configuration uses a 2,048-token anchor vocabulary, up to 2,048 valid response-token contexts per domain, and a rank-16 truncated SVD. Anchor coordinates are fixed before factorization, with duplicate token IDs removed.

For domain d, let

$$r_{d,t}^{(0)} = (\log q_d(\cdot \mid x_t) - \log p_{\theta_0}(\cdot \mid x_t))_{\mathcal{V}_d}. \tag{63}$$

The matrix used for mode construction contains the clipped residuals

$$\bar{r}_{d,t}^{(0)} = \mathrm{clip}\left(r_{d,t}^{(0)}, -20, 20\right). \tag{64}$$

Truncated SVD gives orthonormal directions $\{v_{d,k}\}_{k=1}^{K_d}$. Their calibration coefficients and energies are

$$a_{d,t,k}^{\mathrm{cal}} = v_{d,k}^{\top} \bar{r}_{d,t}^{(0)}, \qquad E_{d,k}^{\mathrm{cal}} = \frac{1}{N_d^{\mathrm{cal}}} \sum_t \left(a_{d,t,k}^{\mathrm{cal}}\right)^2. \tag{65}$$

Within-domain energy shares are

$$e_{d,k} = \frac{E_{d,k}^{\mathrm{cal}}}{\sum_j E_{d,j}^{\mathrm{cal}}}$$

for nonzero-energy banks. The reconstruction identities in Appendix A.4 apply to this clipped calibration matrix.

Cross-sample reliability. Calibration contexts are partitioned into disjoint splits $\mathcal{C}_d^A$ and $\mathcal{C}_d^B$. Let V_d^A and V_d^B be orthonormal bases constructed from their clipped residual matrices. For each full-bank direction, define

$$p_{d,k}^s = \|(V_d^s)^{\top} v_{d,k}\|_2, \qquad E_{d,k}^s = \frac{1}{|\mathcal{C}_d^s|} \sum_{x_t \in \mathcal{C}_d^s} \left[v_{d,k}^{\top} \bar{r}_{d,t}^{(0)}\right]^2, \quad s \in \{A, B\}. \tag{66}$$

The reliability score is

$$\rho_{d,k} = \left[p_{d,k}^A p_{d,k}^B \frac{\min(E_{d,k}^A, E_{d,k}^B) + \varepsilon_\rho}{\max(E_{d,k}^A, E_{d,k}^B) + \varepsilon_\rho}\right]^{1/2}. \tag{67}$$

The full-bank modes and split subspaces share calibration data, so $\rho_{d,k}$ measures internal consistency in direction and energy. Reliability-adjusted energy provides the component value used in selection.

Coverage and selection settings. Selection uses the coverage initialization and global stage in Appendix B.4. The eligibility test is the configured reliability threshold. Coverage additions use the current set-dependent scores without a nominal-budget rejection.

The source bank is shared across selection variants. Calibration statistics, mode directions, and the final selection mask remain fixed during training.

C.3 Response Estimation and Projection

Probe responses and progress targets. Response estimation uses fixed probe contexts at θ_0. The reported configuration uses four probe examples and at most 256 valid response-token contexts per domain. For mode $i = (d, k)$, the probe coefficient is computed from the unclipped residual:

$$a_{t,i}^{\mathrm{probe}} = v_{d,k}^{\top} r_{d,t}(\theta_0), \qquad x_t \in \mathcal{P}_d. \tag{68}$$

Holding these coefficients fixed, define

$$\mathcal{J}_i(\theta) = \frac{1}{|\mathcal{P}_d|} \sum_{x_t \in \mathcal{P}_d} a_{t,i}^{\mathrm{probe}} v_{d,k}^{\top} \big(\log p_\theta(\cdot \mid x_t)\big)_{\mathcal{V}_d}, \tag{69}$$

$$\delta_i = \frac{1}{|\mathcal{P}_d|} \sum_{x_t \in \mathcal{P}_d} \Big(a_{t,i}^{\mathrm{probe}}\Big)^2. \tag{70}$$

Replacing the initial student response by the teacher response on these same contexts increases $\mathcal{J}_i$ by exactly δ_i. The target is therefore a probe-set quantity, whereas E_i^{cal} measures clipped calibration energy.

Projected signatures. Gradients are computed over both LoRA factors (A, B) in the attention projections (`q_proj`, `k_proj`, `v_proj`, `o_proj`) and MLP projections (`gate_proj`, `up_proj`, `down_proj`) across all Transformer layers. With θ collecting these parameters, we use the identity metric $M = I$:

$$g_i = \nabla_\theta \mathcal{J}_i(\theta_0), \qquad \widetilde{g}_i = P g_i.$$

All candidates share a 1,024-dimensional CountSketch with a fixed seed, applied directly to the gradients without optimizer preconditioning or L2 normalization. Initial mode-bank construction uses Kaiming-initialized A and zero-initialized B. Hence $\nabla_A \mathcal{J}_i(\theta_0) = 0$, so only B-parameter gradients contribute to the initial projected signatures, while the sketch covers both factors. Gradient accumulation is performed in chunks to avoid materializing a dense candidate-by-parameter matrix. Projected response matching uses $\eta = 10^{-3}$.

The set cost and marginal increments are computed from the calibrated projected signatures and probe targets.

C.4 Distillation Targets and Training

Current teacher corrections. Training uses the current unclipped residual,

$$r_{d,t}(\theta) = \left(\log q_d(\cdot \mid x_t) - \log p_\theta(\cdot \mid x_t)\right)_{\mathcal{V}_d}.$$

The selected modes reconstruct

$$\widehat{r}_{d,t}(\theta) = \sum_{k=1}^{K_d} z_{d,k} \big(r_{d,t}(\theta)^{\top} v_{d,k}\big) v_{d,k}. \tag{71}$$

The detached target logits are

$$\widetilde{\ell}_{d,t} = \mathrm{sg}\left[\ell_\theta(x_t) + \mathcal{E}_d(\widehat{r}_{d,t}(\theta))\right]. \tag{72}$$

The embedding assigns zero additional logit correction outside the anchor vocabulary. Within the anchor vocabulary, only the selected span contributes to the correction. CE supervision remains unfiltered.

Support-conditioned distillation. For token t in sample b, let

$$A_{bt} = \mathrm{TopK}_{50}(\ell_{\theta,bt}/T) \cup \mathrm{TopK}_{50}(\widetilde{\ell}_{d,bt}/T) \cup \{y_{bt}\}. \tag{73}$$

Both distributions are normalized within this support:

$$p_{bt}^{A}(j) = \frac{\exp(\ell_{\theta,bt,j}/T)}{\sum_{u \in A_{bt}} \exp(\ell_{\theta,bt,u}/T)}, \qquad \widetilde{q}_{bt}^{A}(j) = \frac{\exp(\widetilde{\ell}_{d,bt,j}/T)}{\sum_{u \in A_{bt}} \exp(\widetilde{\ell}_{d,bt,u}/T)}. \tag{74}$$

For valid response positions $\mathcal{T}_b$, the implemented KD loss is

$$\mathcal{L}_{\mathrm{KD}} = \frac{1}{N_{\mathrm{batch}}} \sum_{b=1}^{N_{\mathrm{batch}}} \frac{T^2}{|\mathcal{T}_b|} \sum_{t \in \mathcal{T}_b} \mathrm{KL}\left(\widetilde{q}_{bt}^{A} \,\|\, p_{bt}^{A}\right). \tag{75}$$

The target and support indices are detached. Probability mass outside A_{bt} is omitted; the loss compares the conditional distributions on the retained support.

Optimization. The reported training configuration uses $T = 1$ and

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + 0.1\,\mathcal{L}_{\mathrm{KD}}.$$

Students are adapted with LoRA Hu et al. (2021) rank 64 and scaling parameter 128 in bfloat16. Training uses a learning rate of 2×10^{-5}, a 3% warmup, and at most 4,500 optimizer updates. Four distributed workers use a per-worker batch size of two and two gradient-accumulation steps. Prompt and response lengths are each capped at 2,048 tokens.

Validation and checkpoint saving occur every 100 updates, and the checkpoint with the lowest validation loss is selected.

C.5 Baselines and Controlled Comparisons

Baseline implementations. We compare with SFT, SoftKD, WeightedKD, MiniLLM, GKD, and AdaKD. PCGrad and reference-KL regularization provide additional controls. Within each family, comparisons share the teacher–student configuration, source datasets, and main-experiment domain assignments. Method-specific objectives and sampling procedures follow the corresponding baseline implementations.

The SoftKD implementation and launch script use the same support-construction and normalization rule as Equation 75. SoftKD uses the original routed teacher target, whereas JFMD uses the detached filtered target. Both implementations average token losses within each sample before averaging samples.

Matched-count selection controls. Mode-selection variants share the source bank, target-construction procedure, CE/KD weights, and downstream training. Each variant matches JFMD's retained mode count separately within each domain.

The separable selector retains the reliability threshold and ranks eligible modes by $\widehat{w}_i / C_i^{\mathrm{sep}}$, where

$$a_i = \frac{D_{d(i)}}{\max(s_i, \varepsilon_{\mathrm{num}})}, \qquad C_i^{\mathrm{sep}} = \frac{a_i}{\max\left(|\mathcal{I}|^{-1} \sum_{j \in \mathcal{I}} a_j, \varepsilon_{\mathrm{num}}\right)}. \tag{76}$$

The domain factor D_d is read from the source mode bank at reselection. During bank construction, it is loaded from `components[d].conditional_pair_cost` in the specified JSON file, or set to one if no file is supplied. To compute the sensitivity s_i, let $\mathcal{T}_d$ be the calibration token contexts used to fit the source mode bank and $\mathcal{V}_d$ its anchor vocabulary. First set

$$\bar{p}_{d,v} = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} p_{\theta_0}(v \mid x_t), \qquad v \in \mathcal{V}_d, \tag{77}$$

and then compute

$$s_i = \max\left\{ \sum_{v \in \mathcal{V}_d} \bar{p}_{d,v} v_{i,v}^2 - \left(\sum_{v \in \mathcal{V}_d} \bar{p}_{d,v} v_{i,v} \right)^2, \varepsilon_{\mathrm{num}} \right\}. \tag{78}$$

The probabilities are normalized over the full vocabulary before anchor coordinates are extracted; they are not renormalized on $\mathcal{V}_d$. The mean probabilities are computed before the quadratic form. The cost mean is computed over all candidate modes before reliability filtering, and C_i^{sep} remains fixed as selection proceeds. Singleton-cost selection replaces conditional costs with individual costs. Value-only selection ranks by reliability-adjusted energy without cost. Energy-only selection removes reliability weighting and cost from the ranking, while random selection discards ranking information. These comparisons vary the retained directions at fixed per-domain mode count.

Routing-error protocol. The routing-error experiment uses Qwen3-1.7B. With probability $r \in \{0, 0.1, 0.2, 0.3\}$, the correct route is replaced by a uniformly sampled incorrect domain. SoftKD and JFMD use the same corruption rate and routing seed.

The corrupted assignment selects the teacher and, for JFMD, the corresponding domain modes and anchor vocabulary. The clean calibration bank is held fixed. After corruption, JFMD selects its correction using the corrupted route.

C.6 Evaluation and Computational Cost

Metrics and aggregation. We report accuracy for mathematics and science, BLEU for translation, and ROUGE-L for summarization. For each seed, the four accuracy scores and four generation scores are averaged separately with equal benchmark weights. The generation aggregate combines BLEU and ROUGE-L and is interpreted alongside individual task results.

Let A_s denote an aggregate for seed s. For n seeds, the reported mean and sample standard deviation are

$$\overline{A} = \frac{1}{n}\sum_{s=1}^{n} A_s, \qquad s_A = \sqrt{\frac{1}{n-1}\sum_{s=1}^{n}(A_s - \overline{A})^2}.$$

Per-task entries are seed means. Aggregate uncertainty is calculated from per-seed aggregates.

For generation benchmarks, decoding settings and maximum output lengths are fixed across methods within each model family. The evaluation batch size is two, with at most 512 newly generated tokens.

Computation. The reported training setup uses four A800 GPUs. JFMD wall-clock times include calibration and mode-bank construction, including response estimation and selection, as well as training. These preprocessing stages are performed once per run.

For k_d selected modes in domain d, projection and reconstruction require $O(|\mathcal{V}_d|k_d)$ arithmetic per supervised token, in addition to teacher and student evaluation. Teachers, response sketches, and the mode bank are unnecessary at inference. With the same LoRA deployment convention, the trained student uses the same inference architecture as its baseline counterpart.

D Extended Results and Robustness

This section reports task-level ablations, capacity and routing results, and sensitivity to selection settings.

D.1 Task-Level Results of Controlled Ablations

Table 8 expands the aggregate ablations in Section 5.3. Selection variants share the bank and downstream training and match JFMD's per-domain retained counts. The table reports separable- and singleton-cost selection alongside value-only, energy-only, and random selection.

Task-level patterns. Qwen3's selection gains are concentrated in science: JFMD scores 66.27 on SciQ and 76.67 on OBQA, compared with 58.21 and 61.83 for value-only selection. Singleton-cost selection scores higher on MATH and lower on both science tasks. For LLaMA, random selection scores highest on FLORES, while JFMD scores higher on IWSLT and all four accuracy tasks. For Gemma, JFMD improves all eight tasks over singleton-cost selection.

Within the cost-free controls, reliability weighting improves Qwen3 accuracy but lowers its generation score. Energy-only selection yields higher aggregate means on LLaMA and a slightly higher generation score on Gemma.

D.2 Transfer across Student Capacities

Table 9 reports Qwen3 students with 0.6B, 1.7B, and 4B parameters under fixed domain teachers, expanding Figure 5(a).

Table 8: Complete task-level control and ablation results. Mode-selection variants share the source bank and downstream training, with retained counts matched to JFMD within each domain. The *w/o Reliability & Cost* variant uses energy-only selection. Task columns report means; aggregate columns report mean ± standard deviation.

	Math		Science		Translation		Summarization		Average	
Variant	GSM8K	MATH	SciQ	OBQA	FLORES	IWSLT	XSum	SAMSum	Acc.	Gen.
	Acc.	Acc.	Acc.	Acc.	BLEU	BLEU	R-L	R-L	Avg.	Avg.
Qwen3: 8B → 1.7B										
SoftKD	68.02	62.99	65.52	72.52	26.85	14.72	27.22	45.50	67.26 ± 0.33	28.57 ± 0.38
w/ PCGrad	69.41	64.03	51.10	69.81	26.64	15.48	26.24	44.99	63.59 ± 0.44	28.34 ± 0.21
w/ Reference KL	67.00	65.07	67.55	72.58	26.59	14.76	27.28	45.53	68.05 ± 0.69	28.54 ± 0.59
JFMD	72.86	65.25	66.27	76.67	27.52	17.73	28.20	45.95	70.26 ± 0.15	29.85 ± 0.29
w/o Cost	73.21	66.78	58.21	61.83	27.33	15.66	27.97	45.48	65.01 ± 1.85	29.11 ± 0.23
w/o Reliability & Cost	72.13	64.31	56.01	60.98	28.27	15.76	28.04	45.30	63.36 ± 1.19	29.34 ± 0.50
w/ Random Selector	72.58	62.12	54.80	68.19	27.69	16.24	28.02	45.61	64.42 ± 1.18	29.39 ± 0.23
w/ Separable Selector	72.81	67.84	60.83	70.65	26.93	16.95	27.95	45.88	68.03 ± 1.85	29.43 ± 0.39
w/ Singleton-Cost	71.52	65.44	66.08	71.01	26.63	16.26	27.89	45.69	68.51 ± 0.66	29.12 ± 0.32
LLaMA: 8B → 1B										
SoftKD	13.18	6.11	57.92	46.16	14.84	14.43	28.10	43.36	30.84 ± 0.19	25.18 ± 0.22
w/ PCGrad	13.44	6.80	55.50	49.02	14.33	13.39	29.04	44.30	31.19 ± 0.65	25.26 ± 0.58
w/ Reference KL	13.97	6.77	59.85	49.81	15.30	14.75	28.83	45.04	32.60 ± 0.45	25.98 ± 0.10
JFMD	15.05	7.31	60.87	55.18	17.97	14.91	29.66	44.73	34.60 ± 0.51	26.82 ± 0.18
w/o Cost	15.13	6.53	60.87	54.67	18.00	13.39	29.65	44.46	34.30 ± 0.28	26.37 ± 0.38
w/o Reliability & Cost	15.07	6.73	61.27	54.67	18.42	13.90	29.66	44.41	34.43 ± 0.23	26.60 ± 0.63
w/ Random Selector	14.47	6.67	59.73	53.53	19.40	13.05	29.68	44.53	33.60 ± 0.52	26.67 ± 0.30
w/ Separable Selector	14.53	6.40	61.33	54.27	17.42	13.84	29.58	44.55	34.13 ± 0.40	26.35 ± 0.29
w/ Singleton-Cost	14.35	7.27	59.99	53.89	16.62	13.94	29.44	44.38	33.88 ± 0.33	26.10 ± 0.13
Gemma: 7B → 2B										
SoftKD	33.13	19.76	67.90	74.01	28.58	24.13	32.50	45.83	48.70 ± 0.21	32.76 ± 0.42
w/ PCGrad	29.38	19.75	62.14	68.22	27.56	23.44	31.10	45.64	44.87 ± 0.38	31.93 ± 0.11
w/ Reference KL	30.46	19.80	67.55	74.30	27.67	23.24	32.07	45.90	48.03 ± 0.18	32.22 ± 0.02
JFMD	33.40	21.96	68.87	75.87	29.36	24.68	33.44	46.68	50.02 ± 0.23	33.54 ± 0.17
w/o Cost	32.56	20.94	67.91	75.33	29.24	24.21	33.03	46.94	49.19 ± 0.22	33.36 ± 0.18
w/o Reliability & Cost	32.40	20.94	67.96	75.44	29.24	24.35	33.14	46.86	49.19 ± 0.22	33.40 ± 0.15
w/ Random Selector	32.48	20.83	67.02	73.64	28.56	24.43	32.88	46.81	48.49 ± 0.35	33.17 ± 0.35
w/ Separable Selector	31.98	21.65	68.06	75.40	28.65	24.38	33.19	46.66	49.27 ± 0.22	33.22 ± 0.13
w/ Singleton-Cost	32.35	21.36	68.43	75.01	28.49	23.99	33.15	46.33	49.29 ± 0.17	32.99 ± 0.36

Capacity-dependent gains. In Table 9, JFMD achieves higher mean aggregate scores than SoftKD at all three tested student sizes. The accuracy gains are 1.29, 3.00, and 1.15 points for 0.6B, 1.7B, and 4B students, respectively, while the generation gains are 0.99, 1.28, and 0.42 points. Both aggregate gains are largest at 1.7B, where JFMD also achieves higher mean scores on all eight tasks. The task-level pattern is less uniform at the other sizes: JFMD scores slightly lower on IWSLT at 0.6B and on GSM8K and MATH at 4B. Thus, the aggregate improvements across the tested sizes do not imply improvements on every task.

D.3 Robustness to Teacher Routing Errors

The capacity experiments vary the student while keeping teacher assignments correct. We now examine changes in the supervision itself: incorrect routing exposes the student to a teacher whose specialization does not match the example's domain. Using the Qwen3-1.7B student, we compare SoftKD and JFMD at routing-error rates $r \in \{0, 0.1, 0.2, 0.3\}$. Table 10 reports the aggregate

Table 9: Task-level results across Qwen3 student capacities with fixed domain teachers. Student denotes the corresponding initialization. Task columns report means; aggregate columns report mean ± standard deviation for trained methods. Acc. Avg. averages the four accuracy tasks, and Gen. Avg. averages the four generation scores.

	Math		Science		Translation		Summarization		Average	
Method	GSM8K Acc.	MATH Acc.	SciQ Acc.	OBQA Acc.	FLORES BLEU	IWSLT BLEU	XSum R-L	SAMSum R-L	Acc. Avg.	Gen. Avg.
Qwen3-0.6B										
Student	54.64	35.48	58.36	57.66	21.92	10.34	13.28	29.83	51.54	18.84
SFT	58.20	50.83	60.86	61.24	23.80	11.43	24.68	41.48	57.78 ± 0.58	25.35 ± 0.35
SoftKD	58.81	55.61	61.68	60.86	21.98	13.75	25.24	43.62	59.24 ± 0.48	26.15 ± 0.28
JFMD	59.22	56.88	63.24	62.77	24.57	13.62	25.88	44.49	60.53 ± 0.51	27.14 ± 0.33
Qwen3-1.7B										
Student	61.43	42.22	64.88	70.00	25.27	15.48	12.21	29.35	59.63	20.58
SFT	67.65	62.71	65.31	72.00	25.85	15.85	27.07	45.27	66.92 ± 0.56	28.51 ± 0.24
SoftKD	68.02	62.99	65.52	72.52	26.85	14.72	27.22	45.50	67.26 ± 0.33	28.57 ± 0.38
JFMD	72.86	65.25	66.27	76.67	27.52	17.73	28.20	45.95	70.26 ± 0.15	29.85 ± 0.29
Qwen3-4B										
Student	74.88	57.42	70.48	79.63	26.70	22.21	15.34	29.92	70.60	23.54
SFT	78.68	68.79	68.48	79.33	29.07	22.79	28.42	46.51	73.82 ± 0.41	31.70 ± 0.33
SoftKD	78.88	67.92	71.44	81.72	30.78	22.74	30.87	46.68	74.99 ± 0.35	32.77 ± 0.21
JFMD	78.42	67.80	74.38	83.96	31.19	23.44	31.22	46.91	76.14 ± 0.45	33.19 ± 0.16

Table 10: Performance under teacher routing errors for Qwen3-1.7B. Entries show aggregate means ± reported errors. Acc. Avg. averages the four accuracy benchmarks, and Gen. Avg. averages the four generation scores.

	SoftKD		JFMD	
Routing error	Acc. Avg.	Gen. Avg.	Acc. Avg.	Gen. Avg.
0%	67.26 ± 0.33	28.57 ± 0.38	70.26 ± 0.15	29.85 ± 0.29
10%	58.60 ± 0.23	28.32 ± 0.48	69.81 ± 0.41	29.55 ± 0.35
20%	56.00 ± 0.44	28.12 ± 0.37	69.42 ± 0.35	29.25 ± 0.28
30%	52.65 ± 0.31	27.82 ± 0.32	68.71 ± 0.52	29.30 ± 0.33

results corresponding to Figure 5(b). We then examine component-level interventions to compare their effects on target quality and gradient alignment.

Routing sensitivity. At 30% routing error, SoftKD and JFMD lose 14.61 and 1.55 accuracy points, respectively (Table 10). Their generation losses are 0.75 and 0.55 points; the robustness difference is concentrated in accuracy.

Component interventions and aggregation. To examine the supervision underlying this robustness difference, we separate routing-error corrections into the components retained and removed by JFMD's mode filter. We compare the correctly routed JFMD target with interventions introducing the retained, removed, or full error component. The diagnostics measure ground-truth answer NLL under the resulting target distributions and alignment between the KD and supervised gradients. These diagnostics characterize the immediate effects of target interventions.

The analysis uses five seeds. For each seed, it includes 128 examples per domain and evaluates all three incorrect teacher assignments for each example, with equal weight. Within each seed and domain, target NLL is computed by dividing the accumulated NLL sum by the accumulated context weight. Gradient alignment is computed from the accumulated gradient inner products and squared

Table 11: Component interventions under teacher routing errors. ΔNLL is the change in ground-truth answer NLL relative to the correctly routed JFMD target, in nats per token. Gradient cosine measures alignment between the KD and supervised gradients. Entries are domain-balanced means ± sample standard deviations across five seeds.

Target intervention	ΔNLL	Gradient cosine
Correct routing	0	$+0.0894 \pm 0.0020$
Retained error	-0.393 ± 0.014	$+0.0729 \pm 0.0033$
Removed error	$+3.589 \pm 0.050$	-0.0473 ± 0.0008
Full error	$+2.924 \pm 0.019$	-0.0768 ± 0.0005

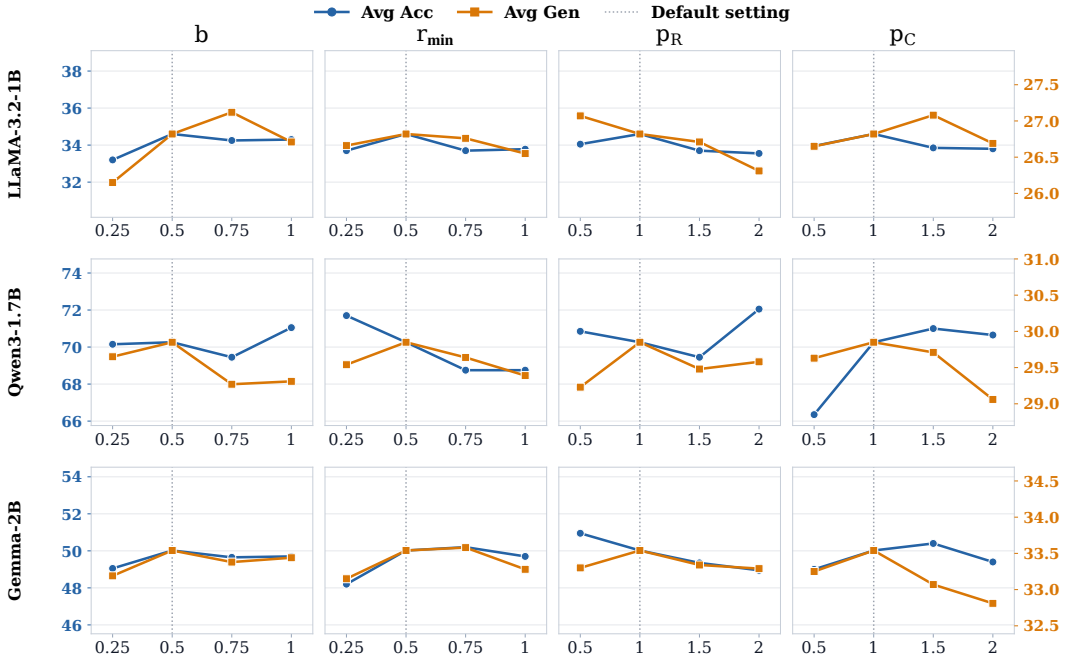


Figure 8: Sensitivity to selection settings across the LLaMA (top), Qwen3 (middle), and Gemma (bottom) student configurations. Columns vary the budget control b, reliability threshold $r_{\min}$, reliability exponent p_R, and cost exponent p_C, respectively. Each panel changes one setting while holding the others fixed. Dotted vertical lines indicate the reference settings $(0.5, 0.5, 1, 1)$. Blue circles show average accuracy over the four accuracy benchmarks; orange squares show the average generation score over the four generation benchmarks.

norms:

$$c_{s,d} = \frac{\sum_j \langle g_j^{\mathrm{KD}}, g_j^{\mathrm{sup}} \rangle}{\sqrt{\left(\sum_j \|g_j^{\mathrm{KD}}\|_2^2\right)\left(\sum_j \|g_j^{\mathrm{sup}}\|_2^2\right)}},$$

where j indexes the diagnostic records within seed s and domain d. We average the resulting statistics equally across the four domains within each seed, then report the mean and sample standard deviation across seeds. This domain-balanced aggregation is used throughout the component analysis.

Component effects. Table 11 associates removed-error corrections with higher answer NLL and negative KD–supervised gradient alignment. Retained-error corrections lower NLL on average and preserve positive alignment. These signs hold across all five domain-balanced seed estimates. Target normalization and logarithmic loss make the combined NLL effect nonlinear.

Energy shares are computed within each seed and domain as accumulated component energy divided by accumulated full-error energy, then averaged with the same domain-balanced rule. The removed component contains 9.58% of the error energy and the retained component 90.42%. The lower-energy removed-error intervention produces the larger NLL increase.

Domain-level patterns. Removed-error interventions increase mean NLL in all four domains, most strongly in science. Retained-error interventions lower NLL in mathematics, science, and translation and slightly raise it in summarization. The removed component's translation-domain gradient cosine is positive, while its domain-balanced mean is negative.

D.4 Sensitivity to Selection Settings

Figure 8 reports one-at-a-time selection-setting sweeps for the three model configurations.

Perturbations around the selection rule. We vary the budget control b and the reliability threshold $r_{\min}$, which determine the available cost budget and candidate eligibility. We also examine changes in the relative influence of reliability and marginal cost through

$$\mathrm{Score}_{p_R,p_C}(i \mid S) = \frac{\rho_i^{p_R} e_i}{\left(\widehat{\Delta C}_{i|S} + \varepsilon\right)^{p_C}}. \tag{79}$$

The setting $p_R = p_C = 1$ recovers the selection rule in Equation 11. The exponent sweeps therefore examine departures from the fixed reliability-adjusted value-per-cost rule used in the main experiments.

Each panel varies one setting while holding the others at $(b, r_{\min}, p_R, p_C) = (0.5, 0.5, 1, 1)$. The dotted vertical lines mark these reference settings.

Sensitivity across model families. LLaMA and Gemma show relatively small changes in both aggregates. Qwen3 accuracy is more sensitive to the reliability threshold and score exponents, while its generation average varies over a narrower range.