X-OPSD: PRESERVING CULTURAL SPECIFICITY IN CROSS-LINGUAL REASONING TRANSFER

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-lingual reasoning transfer can strengthen multilingual large language models (LLMs) while disrupting culturally grounded predictions. We identify this failure mode as *transfer spillover* and hypothesize that representation changes supporting stronger reasoning can be partially decoupled from those interfering with cultural specificity. To investigate this, we introduce **X-OPSD**, a culture-preserving on-policy self-distillation framework that constructs complementary implicit teachers from a single frozen backbone. Reasoning-transfer teachers attenuate language-related activation components to elicit stronger reasoning, while culture-preserving teachers restrict these updates to directions outside an identified culture-related subspace. Their task-dependent supervision is distilled into a single student that requires neither activation steering nor teacher routing at inference time. Across three state-of-the-art multilingual backbones, X-OPSD improves mathematical reasoning and culture-sensitive benchmark averages over the base models, while reducing cultural forgetting relative to English-privileged distillation. Representation analyses and teacher ablations support the role of selective intervention in mitigating spillover. These findings motivate making cultural preservation part of how reasoning supervision is constructed, rather than treating it as a separate objective after transfer.

1 INTRODUCTION

Cross-lingual transfer offers a practical route to improving multilingual reasoning by reusing capabilities that are already stronger in high-resource languages. Early approaches translate questions or supervision across languages, while more recent methods align multilingual reasoning trajectories or use English-privileged teachers to supervise target-language generation (Li et al., 2023; Shi et al., 2022; Zhu et al., 2024; Zhang et al., 2024; Liu et al., 2026; Wang et al., 2026). These methods share a useful premise: a capability learned well in one language need not be relearned independently in every other language. For mathematical reasoning, where the underlying procedure is largely reusable, this premise is particularly effective. The harder question is what else is transferred along with that capability.

Language-specific information is not only a surface-form burden. It also carries lexical distinctions, regional facts, institutions, conventions, and culturally situated values that may be necessary for answering a question correctly (Romanou et al., 2024; Myung et al., 2024; Naous et al., 2024; Hu et al., 2025). Consequently, a transfer mechanism that makes a model reason more like its English-privileged counterpart can improve reusable reasoning while simultaneously disturbing information that should remain context dependent. We observe precisely this tension in Figure 1: after English-privileged transfer, a model that originally selects the correct German term *Altwassersee* instead follows a misleading association with English *oxbow* and changes its answer to *Biegungssee*. We call such losses **transfer spillover**: reasoning-oriented transfer disrupts predictions that were previously correct or contextually appropriate. Importantly, aggregate accuracy alone can hide this behavior, because newly solved questions may offset previously correct answers that are lost. This motivates our central question: *how can cross-lingual transfer strengthen reusable reasoning without overwriting culturally grounded distinctions?*

Representation-space studies provide a useful way to approach this question. Multilingual LLMs exhibit both shared and language-sensitive structure, and internal computation can show a strong

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.

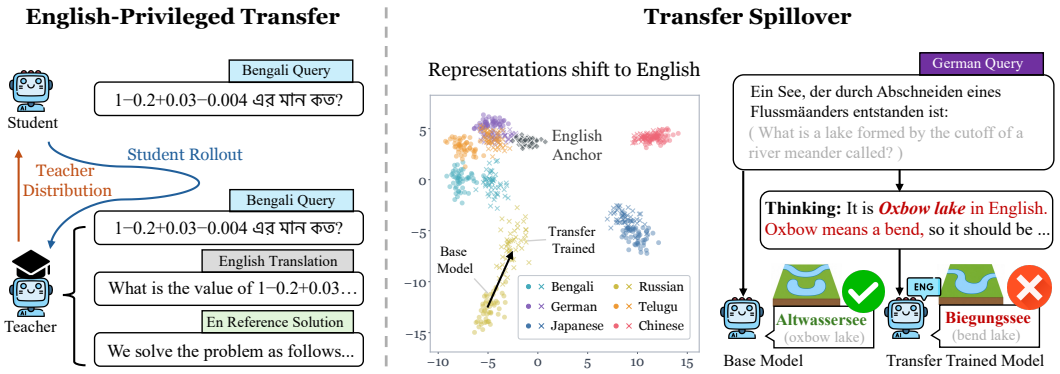


Figure 1: **English-privileged transfer and transfer spillover.** Left: COPSD-style supervision uses English translations and reference solutions. Middle: representations suggest stronger English anchoring after transfer. Right: a misleading English lexical association displaces a previously correct German answer.

English bias even for non-English inputs (Chang et al., 2022; Wendler et al., 2024). MRRE further shows that suppressing language-specific activation components can improve multilingual reasoning (Zhao et al., 2025). However, the same representation change need not be desirable for every input. Moving a mathematical problem toward a shared, English-aligned region can expose stronger reasoning, whereas applying the same change to a question whose answer depends on regional or language-specific knowledge can remove useful distinctions. Our hypothesis is therefore not that language-specific structure should be globally suppressed, but that *representation changes supporting stronger reasoning can be partially decoupled from those that interfere with cultural specificity*. This suggests that preservation should be enforced when the transfer signal is constructed, rather than repaired after unrestricted transfer.

To this end, we introduce **X-OPSD**, a culture-preserving framework for multi-teacher on-policy self-distillation. X-OPSD first uses parallel mathematical inputs to identify language-related directions whose attenuation elicits stronger multilingual reasoning. We then estimate a separate culture-related subspace from culture-sensitive inputs and remove its components from the reasoning intervention, yielding a residual intervention that changes the model while leaving the identified cultural directions untouched. These two intervention spaces instantiate complementary teachers from the same frozen backbone: a *reasoning-transfer teacher* for mathematical inputs and a *culture-preserving teacher* for culture-sensitive inputs. Rather than deploying either intervention at inference time, the student generates its own trajectories and learns from the task-matched teacher's token-level distribution. The resulting student thus consolidates task-dependent transfer into one model, without a larger external teacher, per-example reference solutions, activation steering, or teacher routing at inference time.

We evaluate X-OPSD on three state-of-the-art multilingual backbones using mathematical reasoning and culture-sensitive benchmarks. X-OPSD improves both reasoning and culture-sensitive benchmark-group averages over the corresponding Base models on all three backbones, and it improves both group averages over English-privileged COPSD. Where MRRE is evaluated, X-OPSD also provides stronger joint reasoning and culture-sensitive performance. Paired CulturALL evaluation further shows lower forgetting than COPSD across all three backbones, indicating that the gains reflect better retention of previously correct behavior rather than only additional newly solved examples (cf. Section 4.1 and Appendix J).

We further conduct three analyses to understand *why* the selective teacher construction works. First, to test whether the student internalizes the teachers' different treatment of the two input types rather than merely improving task scores, we analyze held-out representations together with paired forgetting. Mathematical inputs move closer to the English reference, whereas culture-sensitive inputs develop the opposite late-layer trend; this differentiation is accompanied by lower forgetting of previously correct cultural behavior (cf. Section 4.2). Second, to isolate the role of our design choices, we ablate cultural protection and task-dependent teacher assignment. Removing protection weakens culture-sensitive performance, while using a single teacher configuration degrades the overall

transfer–retention balance, showing that the benefit comes from complementary rather than uniform intervention (cf. Section 4.3). Third, we inspect student predictions on fixed mathematical prefixes and find that distillation shifts them toward the reasoning teacher, showing that the teacher's task-appropriate bias remains visible even after steering and routing are removed (cf. Section 4.4). Together, these results support culture-preserving teacher construction as a way to transfer stronger reasoning while retaining distinctions that should remain context dependent.

2 PRELIMINARIES

We briefly review the two ingredients directly used by X-OPSD. We first introduce on-policy self-distillation and its cross-lingual instantiation in COPSD, and then describe representation-based multilingual reasoning transfer through language-subspace intervention, following MRRE. These two mechanisms provide, respectively, the *learning objective* and the *teacher construction interface* used by our method.

2.1 ON-POLICY SELF-DISTILLATION

On-policy distillation (OPD) trains a student on trajectories sampled from the student itself, while a teacher supplies dense next-token supervision on the same prefixes (Agarwal et al., 2024). On-policy self-distillation (OPSD) constructs this teacher from the same model family by giving it privileged information that is unavailable to the student (Zhao et al., 2026). Let p_θ denote the student, let $\hat{y} \sim p_\theta(\cdot \mid x)$ be a student-generated trajectory, and let q_t denote the self-teacher's next-token distribution on prefix $\hat{y}_{<t}$. A trajectory-averaged objective is

$$\mathcal{L}_{\mathrm{OPSD}}(\theta) = \mathbb{E}_{x\sim\mathcal{S},\, \hat{y}\sim p_\theta(\cdot|x)} \left[\frac{1}{|\hat{y}|} \sum_{t=1}^{|\hat{y}|} \mathcal{D}(\mathrm{sg}[q_t], p_\theta(\cdot \mid x, \hat{y}_{<t})) \right], \tag{1}$$

where $\mathcal{D}$ compares the two next-token distributions and sg stops gradients through the teacher target. The rollout is therefore on-policy with respect to the student, while the teaching advantage is determined entirely by how q_t is constructed.

COPSD applies this principle to cross-lingual reasoning (Liu et al., 2026). The student receives the target-language question and generates its own trajectory; the self-teacher scores the same prefixes while additionally observing an English translation of the question and an English reference solution. In this way, stronger English-accessible reasoning can be transferred into target-language generation without exposing the privileged context at inference time. COPSD is therefore a direct example of *behavioral transfer through privileged teacher distributions*: once the teacher is defined, OPSD determines how that behavior is internalized by the student.

2.2 REPRESENTATION-BASED MULTILINGUAL REASONING TRANSFER

A complementary line of work improves multilingual reasoning by intervening on internal language-related representations rather than by providing privileged text. MRRE identifies a low-rank language-specific subspace from language-wise mean activations of parallel inputs (Zhao et al., 2025). At layer k, let $\mathbf{U}_L^{(k)}$ be an orthonormal basis for this subspace and $\mathbf{P}_L^{(k)} = \mathbf{U}_L^{(k)}\mathbf{U}_L^{(k)\top}$ its projector. A basic intervention attenuates the language-related component as

$$\widetilde{\mathbf{h}}^{(k)} = \mathbf{h}^{(k)} - \alpha \mathbf{P}_L^{(k)} \mathbf{h}^{(k)}, \tag{2}$$

where α controls the intervention strength. Applied in appropriate layers, this operation can improve multilingual reasoning by reducing language-dependent variation. MRRE's layer-wise analysis also shows that upper-layer language features are important for maintaining the requested output language, motivating selective rather than uniform intervention across layers.

COPSD and MRRE expose two complementary views of cross-lingual transfer. COPSD shows how an advantage can be distilled from a privileged self-teacher, while MRRE shows how a useful reasoning advantage can be elicited directly from the model's representation space. However, neither mechanism by itself specifies which parts of a reasoning-enhancing representation change should be preserved when the input depends on language- or culture-specific information. X-OPSD connects

the two: we use representation interventions to *construct* self-teachers, constrain those interventions with a culture-related subspace, and then use OPSD to internalize the resulting task-dependent supervision.

3 X-OPSD: CULTURE-PRESERVING ON-POLICY SELF-DISTILLATION

3.1 OVERVIEW

X-OPSD makes two design choices: which representation changes should shape teacher supervision, and which teacher should supervise each task. We first construct reasoning-transfer and culture-preserving teachers from one frozen backbone, then distill their complementary feedback on student-generated trajectories.

3.2 LANGUAGE-CONDITIONED REASONING-TRANSFER TEACHERS

Identifying language-related directions. We instantiate the language-subspace construction from Section 2.2 on parallel mathematical inputs in eight languages. Retaining seven directions at each layer gives the reasoning-related language basis $\mathbf{U}_R^{(k)}$ and projector $\mathbf{P}_R^{(k)} = \mathbf{U}_R^{(k)}\mathbf{U}_R^{(k)\top}$. Appendix D gives the construction and data details.

Constructing reasoning-transfer teachers. For an input in language ℓ, we attenuate the projected component at all token positions in the selected middle layers:

$$\widetilde{\mathbf{h}}_{R,\ell}^{(k)} = \mathbf{h}^{(k)} - \alpha_{R,\ell}\mathbf{P}_R^{(k)}\mathbf{h}^{(k)}. \tag{3}$$

The language-specific coefficient $\alpha_{R,\ell}$ controls attenuation, while components orthogonal to the language-related subspace are preserved. Middle-layer attenuation is followed by upper-layer restoration to retain output-language fidelity. We denote this configured model by the **Reasoning-Transfer Teacher** T_ℓ^R; the layer windows and restoration rule are given in Appendix D.

3.3 CULTURE-PRESERVING TEACHERS

Identifying culture-related directions. To capture variation associated with regional knowledge and culturally specific contexts, we apply the same input-state extraction and subspace construction to the INCLUDE construction split (Romanou et al., 2024). This yields a rank-7 culture-related basis $\mathbf{U}_C^{(k)}$ over the same eight languages. Rather than suppressing these directions, we exclude them from the reasoning-oriented intervention.

Protecting culture-related components. We remove the projection of the mathematical basis onto the culture-related space:

$$\mathbf{U}_\perp^{(k)} = \mathbf{U}_R^{(k)} - \mathbf{U}_C^{(k)}\big(\mathbf{U}_C^{(k)\top}\mathbf{U}_R^{(k)}\big). \tag{4}$$

After orthonormalizing the residual directions, let $\mathbf{P}_\perp^{(k)}$ be the projector onto their span. The **Culture-Preserving Teacher** T_ℓ^C applies

$$\widetilde{\mathbf{h}}_{C,\ell}^{(k)} = \mathbf{h}^{(k)} - \alpha_{C,\ell}\mathbf{P}_\perp^{(k)}\mathbf{h}^{(k)} \tag{5}$$

at all token positions in the selected middle layers. The update is orthogonal to the identified culture-related subspace, so each intervention preserves the current activation's projection onto that subspace. Protection thus constrains the update rather than removing cultural information from the representation.

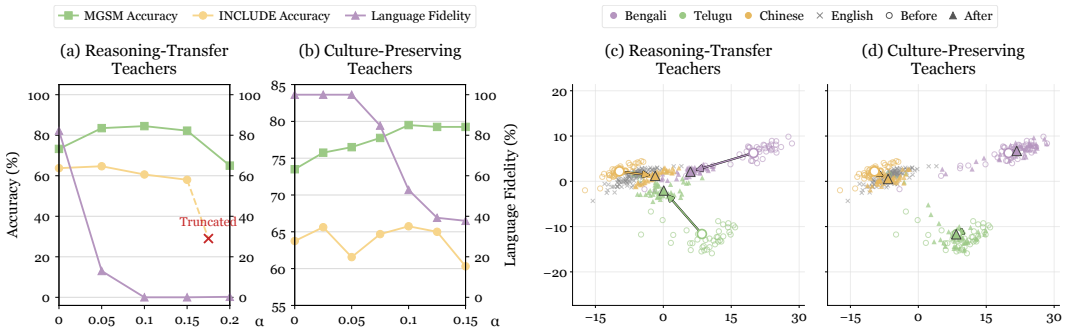


Figure 2: **Effects of the two teacher interventions.** Panels (a) and (b) show MGSM accuracy, INCLUDE accuracy, and language fidelity across intervention strengths for the reasoning-transfer and culture-preserving teachers, respectively. Panels (c) and (d) show the corresponding multilingual representation shifts.

Intervention effects motivate complementary supervision. Figure 2(a,b) shows different behavioral trade-offs. The reasoning-transfer intervention produces larger MGSM gains but stronger declines in INCLUDE accuracy and language fidelity. The culture-preserving intervention yields more moderate reasoning gains while better retaining both. The corresponding representation views in Figure 2(c,d) show larger English-directed shifts under the reasoning intervention and smaller shifts under the protected intervention. Together, these patterns motivate using the reasoning-transfer teacher for mathematical inputs and the culture-preserving teacher for context-sensitive inputs.

3.4 Multi-Teacher On-Policy Self-Distillation

Training combines multilingual MGSM prompts with the INCLUDE construction split. For input x in language $\ell(x)$, the student samples $\hat{y} \sim p_\theta(\cdot \mid x)$ without steering. The task-matched teacher then evaluates the same prefixes:

$$q_t = \begin{cases} T^R_{\ell(x)}(\cdot \mid x, \hat{y}_{<t}), & x \in \mathcal{D}_{\text{MGSM}}, \\ T^C_{\ell(x)}(\cdot \mid x, \hat{y}_{<t}), & x \in \mathcal{D}_{\text{INCLUDE}}. \end{cases} \tag{6}$$

We optimize Eq. 1 with reverse KL, $\mathcal{D}(\text{sg}[q_t], p_{\theta,t}) = D_{\text{KL}}(p_{\theta,t} \,\|\, \text{sg}[q_t])$, where $p_{\theta,t} = p_\theta(\cdot \mid x, \hat{y}_{<t})$. Supervision acts on next-token distributions rather than hidden states. Teacher parameters and language-specific coefficients remain fixed throughout training. Only student parameters are updated, with sampled trajectories treated as fixed in each update. After training, the student operates without steering or teacher routing. Appendix D reports training settings and language-specific intervention coefficients.

4 Experiments

We evaluate reasoning enhancement and cultural retention, then examine three aspects of the learned behavior: **RQ1**, task-dependent representation and retention patterns; **RQ2**, the contributions of protection and teacher assignment; and **RQ3**, reasoning-teacher preference on mathematical prefixes.

4.1 Multilingual Reasoning with Cultural Retention

We evaluate Qwen3.5-4B, Gemma4-E4B, and Qwen3.5-35B-A3B against their Base models, OPD with a larger teacher, OPSD (Zhao et al., 2026), and COPSD (Liu et al., 2026). We also reproduce MRRE (Zhao et al., 2025) on Qwen3.5-4B and Gemma4-E4B using the parallel mathematical data from Language-Reasoning-Disentangle.

We organize evaluation into the same two groups used throughout the paper. The *reasoning* group contains MGSM (Shi et al., 2022), PolyMath (Wang et al., 2025), and an eight-language subset of mAceReason-Math. The *culture-sensitive* group contains INCLUDE (Romanou et al., 2024), CulturALL (Lin et al., 2026), and UniMoral (Kumar & Jurgens, 2025). R-Avg. and C-Avg. average the

Table 1: **Main multilingual results (%).** R-Avg./C-Avg. average the three benchmarks in each group; Lang. Fid. measures response-language agreement. Green/red denote gains/losses relative to Base; bold marks the best score per backbone.

Method	Reasoning				Culture-sensitive				Lang.
	MGSM	PolyMath	mAce	R-Avg.	INCLUDE	CulturALL	UniMoral	C-Avg.	Fid.
Qwen3.5-4B									
Base	62.50	41.11	52.75	52.12	63.38	26.00	40.70	43.36	70.40
OPD	**65.75**	40.07	55.00	53.61	65.00	26.40	38.94	43.45	77.21
OPSD	65.25	42.89	56.00	54.71	65.50	27.20	43.37	45.36	74.47
COPSD	63.63	37.71	47.75	49.70	62.13	28.80	36.44	42.45	66.46
MRRE	44.88	28.46	24.00	32.45	54.13	16.00	33.00	34.38	36.58
X-OPSD	65.63	**46.11**	**62.50**	**58.08**	**65.63**	**29.20**	**46.03**	**46.95**	**81.46**
Gemma4-E4B									
Base	71.63	46.89	36.00	51.51	66.25	28.40	51.80	48.82	95.85
OPD	70.00	45.11	**45.75**	**53.62**	**67.25**	31.60	52.14	50.33	**99.11**
OPSD	70.75	**47.18**	33.50	50.48	66.00	30.80	54.11	50.30	95.54
COPSD	70.25	47.07	35.50	50.94	66.00	28.00	51.79	48.60	95.76
MRRE	71.88	44.39	35.50	50.59	64.88	**32.80**	54.18	50.62	99.08
X-OPSD	**74.88**	46.93	36.50	52.77	**67.25**	30.80	**55.34**	**51.13**	96.82
Qwen3.5-35B-A3B									
Base	70.50	60.25	74.25	68.33	74.88	32.40	44.18	50.48	86.18
OPSD	71.38	**62.11**	73.50	68.99	75.63	**34.40**	**44.72**	**51.58**	85.60
COPSD	71.63	60.11	73.00	68.24	74.88	22.80	43.51	47.06	85.92
X-OPSD	**75.75**	61.86	**77.25**	**71.62**	**76.13**	33.60	44.41	51.38	84.35

three benchmarks in the corresponding group. Language fidelity measures input–output language consistency, averaged over MGSM, PolyMath, and INCLUDE. Table 1 reports the aggregate results; Appendix E provides evaluation settings.

Reasoning gains extend beyond the training benchmark. X-OPSD improves the reasoning average over Base and COPSD on all three backbones and leads the compared methods on both Qwen models. It also exceeds MRRE on both evaluated backbones. The gains extend to mAceReason-Math, which is not used for distillation, showing that the learned behavior generalizes beyond the mathematical training benchmark. Notably, this improvement is obtained while using culture-preserving supervision on a substantial fraction of the training inputs, suggesting that selective teacher routing does not dilute the transferable reasoning signal. Culture-preserving supervision therefore coexists with stronger reasoning rather than simply trading reasoning gains for retention.

Culture-sensitive performance improves without sacrificing retention. X-OPSD improves the cultural average over Base on all three backbones and outperforms COPSD on every culture-sensitive benchmark. Improvements extend to CulturALL, which supplies neither subspace-construction nor training data, indicating that the benefit is not confined to the cultural data used to construct the protected teacher. This is particularly important because aggregate gains on culture-sensitive benchmarks could otherwise arise by solving new examples while overwriting previously correct behavior. We return to paired forgetting in RQ1 to test this distinction directly; Appendix J provides the complete paired results.

Multilingual gains extend across target languages. Figure 3 and Appendix F show reasoning gains across most target languages, alongside broad improvements in culture-sensitive performance. On Qwen3.5-4B, stronger task performance is accompanied by improved language fidelity, while Gemma4-E4B retains high fidelity. The improvements are therefore not driven solely by a small subset of languages or by shifting generation toward a dominant language. The language-level results show that the aggregate gains extend beyond a single target language; Appendix F provides the complete breakdown.

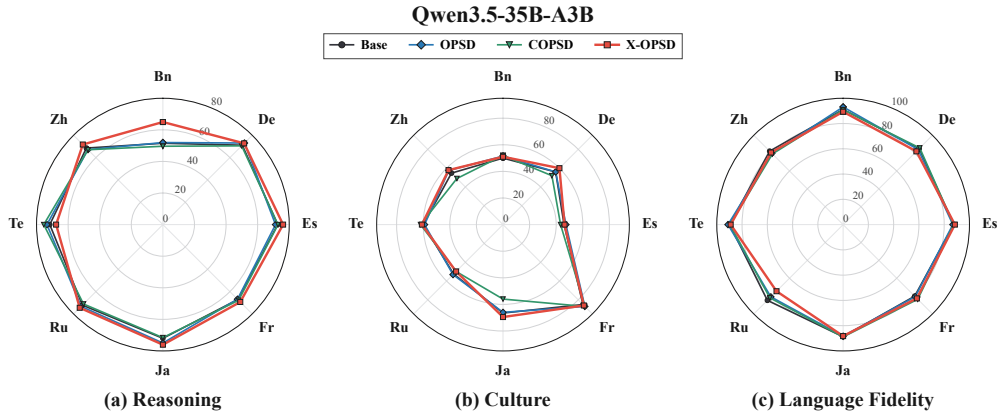


Figure 3: **Language-level results on Qwen3.5-35B-A3B.** Panels show reasoning, culture-sensitive performance, and language fidelity. Each panel uses a common radial scale across languages and methods.

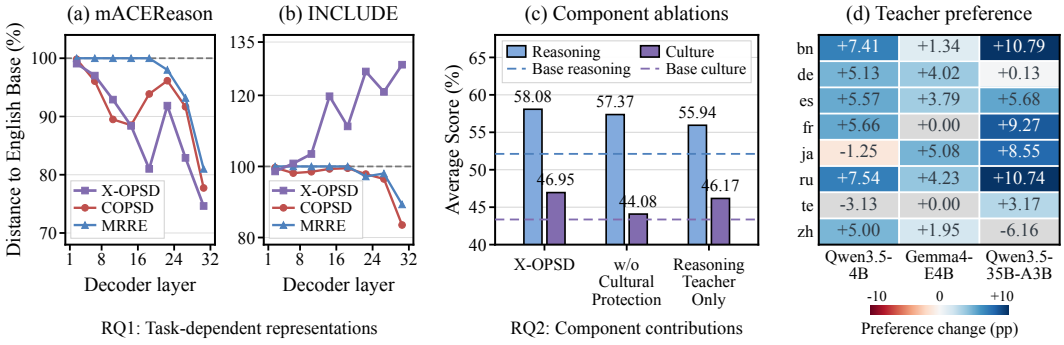


Figure 4: **What does the student learn from complementary teachers? (a,b) RQ1:** English-relative representation distances on held-out mAceReason-Math and INCLUDE inputs for Qwen3.5-4B. Distances use fixed task subspaces and a shared Base English reference; the dashed line is the target-language Base normalization (100). Points average four decoder layers. **(c) RQ2:** Reasoning and culture averages over three benchmarks each, comparing X-OPSD with two teacher variants on Qwen3.5-4B. Dashed lines mark the corresponding Base averages. **(d) RQ3:** Base-adjusted preference for the reasoning teacher over the same-language culture teacher on fixed MGSM prefixes, in percentage points. Gray cells have zero reasoning intervention. Additional results and analysis details appear in Appendices G–I.

4.2 RQ1: Does the Student Learn Different Treatment of Reasoning and Culture?

The main results establish a better reasoning–culture trade-off, but do not show whether the student has actually internalized different processing modes for the two input types. We therefore examine both its internal representations and the behavioral forgetting associated with transfer spillover. A successful student should not merely become uniformly more English-aligned; rather, it should move toward shared representations when that helps reasoning, while retaining separation when the answer depends on context-specific information. Figure 4(a,b) compares X-OPSD, COPSD, and MRRE on held-out mAceReason-Math and INCLUDE inputs. We measure each target-language representation's distance to its paired English Base reference in a fixed task subspace, normalized by the target-language Base distance. Appendix G details the protocol.

Reasoning inputs become more English-aligned. On mAceReason-Math, X-OPSD moves closer to the English reference through the middle and upper layers, with stronger late-layer alignment than COPSD and MRRE. This follows the direction induced by the reasoning-transfer intervention in Figure 2(c). Its appearance on a benchmark not used for distillation, and without inference-time steering, suggests that the student has internalized the teacher's bias toward shared cross-lingual representations. The effect therefore reflects a learned processing tendency rather than a test-time artifact of the intervention.

Culture-sensitive inputs follow a different trajectory. On INCLUDE, X-OPSD instead develops greater separation from the English reference in the upper layers, while COPSD and MRRE remain closer to it. The student receives next-token supervision rather than a hidden-state matching loss, allowing it to realize the teacher's functional objective without reproducing the teacher's intervention trajectory. Together with the cultural gains, this contrast supports task-dependent processing rather than uniform English anchoring: mathematical representations become more aligned, while culture-sensitive representations remain differentiated. This also argues against the simpler explanation that better transfer is driven by a single global shift toward English-like representations.

Lower forgetting confirms that the difference is behavioral, not only geometric. Figure 5 shows that the culture-sensitive representation pattern is accompanied by better retention of previously correct CulturALL answers. X-OPSD consistently reduces forgetting relative to COPSD across the three backbones, while also reducing forgetting relative to MRRE where that comparison is available. Thus, the student does not merely separate culture-sensitive representations from English; it also preserves more of the culturally grounded behavior present in the Base model. The geometric difference is therefore matched by a behavioral reduction in transfer spillover, rather than being only a change in hidden-state organization.

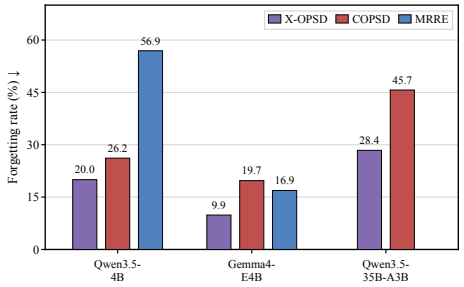


Figure 5: **Paired CulturALL forgetting.** Lower is better; full paired results are in Appendix J.

4.3 RQ2: How Do Protection and Teacher Assignment Affect Transfer?

RQ1 shows that the final student treats reasoning and culture-sensitive inputs differently, but it does not identify which parts of X-OPSD produce this separation. We therefore isolate the two central design choices: protecting culture-related directions and assigning different teachers to the two tasks. These ablations help distinguish whether the benefit comes from the protected geometry itself or from task-dependent teacher routing. We compare two teacher variants on Qwen3.5-4B. *w/o Cultural Protection* replaces the culture route's residual directions with mathematical directions while retaining its coefficients. *Reasoning Teacher Only* instead applies the reasoning teacher, including its intervention configuration, to both mathematical and cultural inputs. The first varies the protection geometry; the second replaces complementary supervision with one teacher configuration. Figure 4(c) compares their reasoning and cultural averages; Appendix H reports benchmark-level results, paired forgetting, and additional variants.

Unprotected supervision weakens cultural retention. Removing protection retains much of the reasoning improvement but lowers the cultural average relative to X-OPSD. INCLUDE and CulturALL fall below Base, and paired forgetting increases. This shows that preserving culture-related directions is not a cosmetic constraint: it materially affects whether transfer retains context-sensitive behavior. The configuration therefore exposes the transfer–retention tension: strong aggregate reasoning can coexist with weaker performance on context-sensitive questions.

Complementarity matters beyond the choice of a single space. Using the reasoning teacher for both tasks lowers both benchmark-group averages and increases forgetting relative to X-OPSD. The additional variants show why neither intervention should be applied uniformly: mathemati-

cal supervision alone loses more previously correct cultural answers, while residual directions on both tasks weaken reasoning and cultural performance. Across these configurations, the complete method offers the strongest joint performance and lowest forgetting, supporting task-specific teacher assignment rather than uniformly stronger protection. This suggests that the gain comes from complementary supervision, not from simply selecting one teacher that is globally best for all inputs.

4.4 RQ3: Does Distillation Increase Reasoning-Teacher Preference?

The representation and ablation analyses show task-dependent behavior, but they do not directly test whether the reasoning teacher leaves a detectable signature in the student's output distribution after routing is removed. We therefore compare student and reasoning-teacher predictions on identical mathematical prefixes from MGSM. Let $O_{10}(A, T)$ denote mean Top-10 next-token overlap. We measure the Base-adjusted preference for the reasoning teacher relative to the same-language culture teacher:

$$\Delta_{\text{route}} = 100\big[O_{10}(S, T_R) - O_{10}(S, T_C) - O_{10}(B, T_R) + O_{10}(B, T_C)\big]. \tag{7}$$

Here, S, B, T_R, and T_C denote the student, Base, reasoning teacher, and culture teacher, respectively. The Base adjustment removes pre-existing similarity to either teacher, so the score isolates the shift introduced by distillation itself. Figure 4(d) shows positive preference changes in almost all language–backbone settings with an active reasoning intervention. Across the three backbones, distillation shifts the relative prediction preference toward the reasoning teacher on mathematical inputs. The supervision's task-appropriate bias thus remains visible in the student's output distribution after steering and explicit routing are removed. Appendix I reports the prefix sampling, language-level preference scores, and confidence intervals.

5 Related Work

Cross-lingual reasoning transfer. Multilingual reasoning research spans translated supervision and cross-lingual prompting (Shi et al., 2022; Qin et al., 2023; Li et al., 2023; Zhu et al., 2024; Zhang et al., 2024), then extends to reasoning-language selection, trajectory alignment, and preference or reinforcement learning (She et al., 2024; Chai et al., 2025; Barua et al., 2026; Son et al., 2026; Huang et al., 2026). Recent on-policy methods refine supervision on student rollouts through privileged context, reasoning-pivot weighting, or teacher–base log-probability differences (Agarwal et al., 2024; Zhao et al., 2026; Liu et al., 2026; Wang et al., 2026; Heo et al., 2026). X-OPSD instead focuses on which representation changes should supply that supervision, coupling reasoning transfer with cultural preservation.

Representation control and cultural preservation. Studies of multilingual geometry, latent language, and language-specific neurons motivate activation-level control (Chang et al., 2022; Wendler et al., 2024; Tang et al., 2024; Kojima et al., 2024), with applications to reasoning and cross-lingual retrieval (Zhao et al., 2025; Sundar et al., 2025; Maraia et al., 2026). Cultural benchmarks and transfer analyses highlight the need to preserve context-dependent distinctions alongside transferable capabilities (Hu et al., 2025; Romanou et al., 2024; Myung et al., 2024; Naous et al., 2024; Han et al., 2025). Surgical Steering addresses this trade-off through layer-specific inference-time interventions (Han et al., 2025); X-OPSD instead constrains teacher construction and distills complementary behavior into an unsteered student.

6 Conclusion

X-OPSD treats preservation as part of cross-lingual teacher construction. Complementary interventions on one frozen backbone supply reasoning and culture-sensitive supervision to an unsteered student. Across three backbones, this design improves both benchmark-group averages, while paired evaluation reveals lower CulturALL forgetting than COPSD. The results argue for evaluating what capability transfer retains, not only what it newly solves. Future work can refine cultural subspaces and extend selective transfer to additional tasks and longer training budgets.

AI USE STATEMENT

Large language models were used to assist with experimental code, analysis, and parts of the experimental pipeline.

ETHICS STATEMENT

This work studies cross-lingual capability transfer with the goal of improving multilingual reasoning while preserving language- and culture-specific information.

REPRODUCIBILITY STATEMENT

Appendices D–J provide implementation details, evaluation settings, language-level results, analysis protocols, ablations, and paired retention results.

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2306.13649.

Josh Barua, Seun Eisape, Kayo Yin, and Alane Suhr. Long chain-of-thought reasoning across languages. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/7dbdf006424e7749c8a35913d3574c4e-Abstract-Conference.html.

Linzheng Chai, Jian Yang, Tao Sun, Hongcheng Guo, Jiaheng Liu, Bing Wang, Xinnian Liang, Jiaqi Bai, Tongliang Li, Qiyao Peng, and Zhoujun Li. xCoT: Cross-lingual instruction tuning for cross-lingual chain-of-thought reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 23550–23558, 2025. doi: 10.1609/aaai.v39i22.34524. URL https://ojs.aaai.org/index.php/AAAI/article/view/34524.

Tyler A. Chang, Zhuowen Tu, and Benjamin K. Bergen. The geometry of multilingual language model representations. *arXiv preprint arXiv:2205.10964*, 2022. URL https://arxiv.org/abs/2205.10964.

HyoJung Han, Sweta Agrawal, and Eleftheria Briakou. Rethinking cross-lingual alignment: Balancing transfer and cultural erasure in multilingual llms. *arXiv preprint arXiv:2510.26024*, 2025. URL https://arxiv.org/abs/2510.26024.

Byeongho Heo, Jaehui Hwang, Sangdoo Yun, and Dongyoon Han. On-policy delta distillation for multilingual math reasoning. *arXiv preprint arXiv:2608.05802*, 2026. URL https://arxiv.org/abs/2608.05802.

Peng Hu, Sizhe Liu, Changjiang Gao, Xin Huang, Xue Han, Junlan Feng, Chao Deng, and Shujian Huang. Large language models are cross-lingual knowledge-free reasoners. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 1525–1542. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.naacl-long.72. URL https://aclanthology.org/2025.naacl-long.72/.

Shulin Huang, Yiran Ding, Junshu Pan, and Yue Zhang. Beyond english-centric training: How reinforcement learning improves cross-lingual reasoning in LLMs. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/9320df227557d00ce68f1b8b07ea2d49-Abstract-Conference.html.

Takeshi Kojima, Itsuki Okimura, Yusuke Iwasawa, Hitomi Yanaka, and Yutaka Matsuo. On the multilingual ability of decoder-based pre-trained language models: Finding and controlling language-specific neurons. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 6919–6971. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.naacl-long.384. URL https://aclanthology.org/2024.naacl-long.384/.

Shivani Kumar and David Jurgens. Are rules meant to be broken? understanding multilingual moral reasoning as a computational pipeline with unimoral. *arXiv preprint arXiv:2502.14083*, 2025. URL https://arxiv.org/abs/2502.14083.

Haonan Li, Fajri Koto, Minghao Wu, Alham Fikri Aji, and Timothy Baldwin. Bactrian-X: Multilingual replicable instruction-following models with low-rank adaptation. *arXiv preprint arXiv:2305.15011*, 2023. URL https://arxiv.org/abs/2305.15011.

Peiqin Lin, Chenyang Lyu, Wenjiang Luo, Haotian Ye, Md Mehrab Hossain, Chunlan Ma, Shaoxiong Ji, Younes Samih, Bo Zeng, Fan Jiang, Yuanbin Cao, Dilda Duisenbek, Adrian Neo Sau Xun, Daria Pozdniakova, Liubou Misevich, Nevena Marinković, Ngoc Gia Linh Nguyen, Thi Khanh Linh Do, Sarakmatak Sophy, Baotian Hu, Guanhua Chen, Gongbo Tang, Alham Fikri

Aji, Longyue Wang, and Weihua Luo. Culturall: Benchmarking multilingual and multicultural competence of llms on grounded tasks. *arXiv preprint arXiv:2604.19262*, 2026. URL https://arxiv.org/abs/2604.19262.

Yihong Liu, Raoyuan Zhao, Michael A. Hedderich, and Hinrich Schütze. Crosslingual on-policy self-distillation for multilingual reasoning. *arXiv preprint arXiv:2605.09548*, 2026. URL https://arxiv.org/abs/2605.09548.

Gabriele Maraia, Leonardo Ranaldi, Marco Valentino, and Fabio Massimo Zanzotto. Can activation steering generalize across languages? a study on syllogistic reasoning in language models. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2739–2753. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.eacl-long.125. URL https://aclanthology.org/2026.eacl-long.125/.

Junho Myung, Nayeon Lee, Yi Zhou, Jiho Jin, Rifki Afina Putri, Dimosthenis Antypas, Hsuvas Borkakoty, Eunsu Kim, Carla Perez-Almendros, Abinew Ali Ayele, Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, Hwaran Lee, Shamsuddeen Hassan Muhammad, Kiwoong Park, Anar Sabuhi Rzayev, Nina White, Seid Muhie Yimam, Mohammad Taher Pilehvar, Nedjma Ousidhoum, Jose Camacho-Collados, and Alice Oh. Blend: A benchmark for llms on everyday knowledge in diverse cultures and languages. *arXiv preprint arXiv:2406.09948*, 2024. URL https://arxiv.org/abs/2406.09948.

Tarek Naous, Michael J Ryan, Alan Ritter, and Wei Xu. Having beer after prayer? measuring cultural bias in large language models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 16366–16393, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.862. URL https://aclanthology.org/2024.acl-long.862/.

Libo Qin, Qiguang Chen, Fuxuan Wei, Shijue Huang, and Wanxiang Che. Cross-lingual prompting: Improving zero-shot chain-of-thought reasoning across languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 2695–2709. Association for Computational Linguistics, 2023. doi: 10.18653/v1/2023.emnlp-main.163. URL https://aclanthology.org/2023.emnlp-main.163/.

Angelika Romanou, Negar Foroutan, Anna Sotnikova, Zeming Chen, Sree Harsha Nelaturu, Shivalika Singh, Rishabh Maheshwary, Micol Altomare, Mohamed A. Haggag, Snegha A, Alfonso Amayuelas, Azril Hafizi Amirudin, Viraat Aryabumi, Danylo Boiko, Michael Chang, Jenny Chim, Gal Cohen, Aditya Kumar Dalmia, Abraham Diress, Sharad Duwal, Daniil Dzenhaliou, Daniel Fernando Erazo Florez, Fabian Farestam, Joseph Marvin Imperial, Shayekh Bin Islam, Perttu Isotalo, Maral Jabbarishiviari, Börje F. Karlsson, Eldar Khalilov, Christopher Klamm, Fajri Koto, Dominik Krzemiński, Gabriel Adriano de Melo, Syrielle Montariol, Yiyang Nan, Joel Niklaus, Jekaterina Novikova, Johan Samir Obando Ceron, Debjit Paul, Esther Ploeger, Jebish Purbey, Swati Rajwal, Selvan Sunitha Ravi, Sara Rydell, Roshan Santhosh, Drishti Sharma, Marjana Prifti Skenduli, Arshia Soltani Moakhar, Bardia Soltani Moakhar, Ran Tamir, Ayush Kumar Tarun, Azmine Toushik Wasi, Thenuka Ovin Weerasinghe, Serhan Yilmaz, Mike Zhang, Imanol Schlag, Marzieh Fadaee, Sara Hooker, and Antoine Bosselut. INCLUDE: Evaluating multilingual language understanding with regional knowledge. *arXiv preprint arXiv:2411.19799*, 2024. URL https://arxiv.org/abs/2411.19799.

Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng, and Jiajun Chen. MAPO: Advancing multilingual reasoning through multilingual-alignment-as-preference optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10015–10027. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.539. URL https://aclanthology.org/2024.acl-long.539/.

Freda Shi, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder, Denny Zhou, Dipanjan Das, and Jason Wei. Language models are multilingual chain-of-thought reasoners. *arXiv preprint arXiv:2210.03057*, 2022. URL https://arxiv.org/abs/2210.03057.

Guijin Son, Donghun Yang, Hitesh Laxmichand Patel, Amit Agarwal, Hyunwoo Ko, Chanuk Lim, Srikant Panda, Minhyuk Kim, Nikunj Drolia, Dasol Choi, Kyong-Ha Lee, and Youngjae Yu. Pushing on multilingual reasoning models with language-mixed chain-of-thought. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/804b5e300c9ed4e3ea3b073f186f4adc-Abstract-Conference.html.

Anirudh Sundar, Sinead Williamson, Katherine Metcalf, Barry-John Theobald, Skyler Seto, and Masha Fedzechkina. Steering into new embedding spaces: Analyzing cross-lingual alignment induced by model interventions in multilingual language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2375–2401. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.acl-long.118. URL https://aclanthology.org/2025.acl-long.118/.

Tianyi Tang, Wenyang Luo, Haoyang Huang, Dongdong Zhang, Xiaolei Wang, Xin Zhao, Furu Wei, and Ji-Rong Wen. Language-specific neurons: The key to multilingual capabilities in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 5701–5715. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.309. URL https://aclanthology.org/2024.acl-long.309/.

Xinye Wang, Junxiao Liu, and Shujian Huang. Rp-opsd: Reasoning-pivot-guided on-policy self-distillation for multilingual reasoning transfer. *arXiv preprint arXiv:2608.06347*, 2026. URL https://arxiv.org/abs/2608.06347.

Yiming Wang, Pei Zhang, Jialong Tang, Haoran Wei, Baosong Yang, Rui Wang, Chenshu Sun, Feitong Sun, Jiran Zhang, Junxuan Wu, Qiqian Cang, Yichang Zhang, Fei Huang, Junyang Lin, Fei Huang, and Jingren Zhou. PolyMath: Evaluating mathematical reasoning in multilingual contexts. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2504.18428.

Chris Wendler, Veniamin Veselovsky, Giovanni Monea, and Robert West. Do llamas work in english? on the latent language of multilingual transformers. *arXiv preprint arXiv:2402.10588*, 2024. URL https://arxiv.org/abs/2402.10588.

Yuanchi Zhang, Yile Wang, Zijun Liu, Shuo Wang, Xiaolong Wang, Peng Li, Maosong Sun, and Yang Liu. Enhancing multilingual capabilities of large language models through self-distillation from resource-rich languages. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 11189–11204. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.603. URL https://aclanthology.org/2024.acl-long.603/.

Siyan Zhao, Zhihui Xie, Mengchen Liu, Jing Huang, Guan Pang, Feiyu Chen, and Aditya Grover. Self-distilled reasoner: On-policy self-distillation for large language models. *arXiv preprint arXiv:2601.18734*, 2026. URL https://arxiv.org/abs/2601.18734.

Weixiang Zhao, Jiahe Guo, Yang Deng, Tongtong Wu, Wenxuan Zhang, Yulin Hu, Xingyu Sui, Yanyan Zhao, Wanxiang Che, Bing Qin, Tat-Seng Chua, and Ting Liu. When less language is more: Language-reasoning disentanglement makes LLMs better multilingual reasoners. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2505.15257.

Wenhao Zhu, Shujian Huang, Fei Yuan, Shuaijie She, Jiajun Chen, and Alexandra Birch. Question translation training for better multilingual reasoning. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 8411–8423. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.498. URL https://aclanthology.org/2024.findings-acl.498/.

A LIMITATIONS

Our work has several limitations that suggest directions for future study. First, the reasoning- and culture-related subspaces are estimated before distillation and remain fixed throughout training. As the student representation evolves, periodically re-estimating or jointly updating these subspaces may enable more precise teacher interventions and further improve the transfer–preservation trade-off. Second, our current construction models reasoning- and culture-related variation using fixed low-rank linear subspaces. While effective in our experiments, culturally relevant information can be context dependent and may not be fully captured by a single linear geometry; future work could explore context-adaptive or nonlinear representations of the protected space. Third, although our evaluation covers multiple model families and scales, including both dense and mixture-of-experts architectures, computational cost limits the breadth of model coverage. In particular, we do not extend experiments to a larger set of backbones or evaluate a substantially larger-teacher OPD baseline for the 35B model. Future work could examine whether the observed transfer–preservation behavior continues to hold across a broader range of model scales and architectures.

B BENCHMARK DESCRIPTIONS

Our six benchmarks cover two complementary aspects of multilingual capability: mathematical reasoning that should transfer across languages, and knowledge and judgments that depend on regional or cultural context. Below we describe each benchmark and the subset used in this study. Generation settings and score aggregation are specified in Appendix E.

MGSM. The Multilingual Grade School Math benchmark (Shi et al., 2022) tests multi-step arithmetic reasoning in natural-language word problems. It was constructed by manually translating 250 GSM8K problems into ten languages, alongside the English originals. Parallel questions preserve the underlying mathematical problem while changing its linguistic form, making MGSM suitable for studying cross-lingual reasoning transfer. Our evaluation uses 100 questions in each of the eight reporting languages, for 800 examples per method. We score the final numerical answer using the fixed response-only parser. MGSM also supplies the mathematical prompts used for X-OPSD training, as detailed in Appendix D.

PolyMath. PolyMath (Wang et al., 2025) evaluates mathematical reasoning across 18 languages and four difficulty levels. Its range of difficulty complements MGSM by testing whether multilingual improvements extend to more demanding problems. Our evaluation uses 150 examples in each of the eight reporting languages, for 1,200 examples per method. We retain the fixed mathematical answer evaluator and within-language difficulty weighting, then average across languages. PolyMath is used for evaluation, not student distillation.

mAceReason-Math. mAceReason-Math (**?**) provides multilingual versions of challenging problems from AceReason-Math, a corpus developed for reinforcement learning with verifiable rewards. The release covers 14 languages and includes parallel training and test splits. We use a fixed subset of its official test split: 50 common integer-answer questions in each of the eight reporting languages, yielding 400 evaluation examples. The subset was sampled with seed 42 from questions whose released answers agree across all eight languages, before model inference. Final numerical answers are scored by exact match. No mAceReason-Math examples are used for student distillation; this benchmark tests transfer beyond the mathematical training source.

INCLUDE. INCLUDE (Romanou et al., 2024) evaluates multilingual understanding using questions collected from local examinations across 44 written languages. It combines knowledge and reasoning with regional content that is not captured by translating an English-only test. We use it to assess performance on locally grounded questions and to construct the culture-preserving teacher. The held-out evaluation subset contains 750 multiple-choice questions: 50 in German and 100 in each of the other seven reporting languages. Performance is measured by answer-choice accuracy. The construction and training subset, coefficient calibration subset, and held-out evaluation subset are disjoint under question-and-option fingerprints (Appendix D).

CulturALL. CulturALL (Lin et al., 2026) tests whether models can apply cultural and regional knowledge within concrete scenarios. Its questions require models to combine contextual cues with relevant knowledge, covering everyday situations as well as region-specific facts and conventions. We evaluate 50 examples each in Bengali, Spanish, Japanese, Russian, and Chinese, for 250 examples per method. Answers are assessed with the frozen DeepSeek Flash non-thinking judge used for the main results. The same questions also support our paired forgetting analysis: we track which Base-correct answers become incorrect after transfer, separately from newly correct answers. CulturALL is not used for student distillation.

UniMoral. UniMoral (Kumar & Jurgens, 2025) examines moral reasoning through annotated dilemmas, action choices, ethical principles, decision factors, and consequences. Its annotations also include participants' moral and cultural profiles. We evaluate four tasks: Action Prediction (AP), Moral Typology Classification (MTC), Factor Attribution Analysis (FAA), and Consequence Generation (CG). AP predicts an action; MTC identifies its ethical category; FAA identifies factors informing the decision; and CG generates consequences of a scenario–action pair. Our evaluation uses 100 examples per language and task in Spanish, Russian, and Chinese. The main-table score averages weighted F1 for AP, MTC, and FAA with BERTScore for CG, following the aggregation in Appendix E. UniMoral is used only for evaluation.

C Descriptions of Compared Methods

The methods in Table 1 differ in how they obtain reasoning supervision and whether intervention is required at inference. OPD, OPSD, COPSD, and X-OPSD learn from student-generated trajectories; MRRE modifies activations during inference.

Base. Base denotes the corresponding released backbone before the additional training or representation interventions studied here: Qwen3.5-4B, Gemma4-E4B, or Qwen3.5-35B-A3B. It receives the evaluation prompt directly and provides the reference for measuring capability gains, language fidelity, and paired cultural retention.

Plain OPD. On-policy distillation (Agarwal et al., 2024) trains a student using feedback from a separate teacher on the student's own generated sequences. At each position, the teacher evaluates the prefix produced by the student and supplies a next-token distribution. Training reduces the discrepancy between student and teacher predictions along these trajectories. Our Plain OPD baseline uses Qwen3.5-35B-A3B as the teacher for Qwen3.5-4B, and Gemma4-26B-A4B as the teacher for Gemma4-E4B. It represents the external-teacher route to improving reasoning, with the trained student deployed on its own.

OPSD. On-policy self-distillation (Zhao et al., 2026) obtains its teaching advantage from privileged information supplied to a self-teacher based on the same backbone. The student receives the question and generates its own trajectory. The teacher also receives a reference answer or solution and evaluates the student's prefixes under this richer context. Token-level distribution matching transfers that advantage to the student, which does not require the privileged information at inference time. This baseline tests reasoning improvement through self-distillation without X-OPSD's task-dependent teacher interventions.

COPSD. Crosslingual on-policy self-distillation (Liu et al., 2026) transfers the model's English-accessible reasoning capability to target languages. The student observes the target-language question, while the self-teacher additionally receives its English translation and an English reference solution. The teacher supervises the same target-language rollout prefixes generated by the student through token-level distribution matching. The trained student answers target-language questions without this English-privileged context at inference time. In our evaluation, COPSD adapters are selected by input language. COPSD provides a direct comparison for English-mediated reasoning transfer without an explicit culture-preserving constraint on teacher construction.

MRRE. MRRE (**?**) enhances multilingual reasoning through representation engineering at inference time. It injects precomputed cross-lingual reasoning vectors to move non-English representations toward English, followed by target-language anchoring vectors to support input–output

language consistency. The backbone weights remain unchanged. We evaluate a project reconstruction on Qwen3.5-4B and Gemma4-E4B, including nonofficial Bengali and Telugu extensions. For Gemma, the intervention is applied during the thinking stage. MRRE therefore compares inference-time representation control with the behavior learned by our unsteered student; its generation settings are recorded in Appendix E.

X-OPSD. Our method constructs complementary teachers through activation interventions on one frozen backbone. Reasoning-transfer teachers attenuate language-related components to expose stronger shared reasoning. Culture-preserving teachers intervene along residual directions obtained by removing components within the culture-related subspace. Mathematical and culture-sensitive training inputs select their respective teacher routes. The student generates unsteered trajectories and learns from the routed next-token distributions using reverse KL, without hidden-state matching or per-example reference solutions. After training, a single student handles both input types without teacher routing or activation steering. Appendix D specifies the data, intervention policies, and optimization settings.

D Implementation Details

Subspace construction. We use the same eight languages throughout: bn, de, es, fr, ja, ru, te, and zh. The mathematical subspace is estimated from 50 parallel prompts per language from the Language-Reasoning-Disentangle mathematical pool (Zhao et al., 2025). The culture-related subspace uses 378 INCLUDE prompts (Romanou et al., 2024); separate INCLUDE splits are used for coefficient calibration and evaluation.

For each prompt, we extract the last input token's activation after each decoder block. At each layer, language-wise means are centered and decomposed with SVD, retaining seven directions to form $\mathbf{U}_R^{(k)}$ or $\mathbf{U}_C^{(k)}$. The culture-preserving space is obtained by residualizing $\mathbf{U}_R^{(k)}$ against $\mathbf{U}_C^{(k)}$ and orthonormalizing the result, as in Eq. 4. The resulting bases are fixed during training and analysis.

Student training. Training combines multilingual MGSM prompts with the INCLUDE construction split. We use full-parameter AdamW optimization with a learning rate of 10^{-5}, a cosine schedule, and bfloat16 precision. Each prompt uses eight sampled completions with temperature 1 and top-p 1. Both task routes use full-vocabulary reverse-KL distillation from the frozen self-teacher.

Training data for baselines. Plain OPD uses a multilingual mixture of 778 prompts: 400 mathematical prompts and 378 INCLUDE prompts. We use Qwen3.5-35B-A3B as the teacher for Qwen3.5-4B, and Gemma4-26B-A4B as the teacher for Gemma4-E4B. The student generates its own responses, and the larger teacher provides token-level supervision on these trajectories.

OPSD uses the official `Openthoughts_math_30k_opsd` training release, containing 29,434 mathematical examples. The student receives the problem, while the self-teacher additionally receives the reference solution as privileged context. We use this training dataset for all three OPSD backbones.

COPSD uses the official `COPSD-PolyMath-TrainDataset` release. We use 3,000 examples per language in Bengali, German, Spanish, French, Japanese, Russian, Telugu, and Chinese. These are eight language versions of the same 3,000 English source problems, yielding 24,000 language-specific records per backbone. Each record contains the target-language problem, the English source problem, and an English reference solution. We train a separate adapter for each target language. The student observes the target-language problem, while the self-teacher additionally receives the English problem and solution to transfer English-accessible reasoning capability.

Base and MRRE require no additional parameter training in our experiments. MRRE applies precomputed representation interventions at inference time.

Intervention layers and strengths. The reasoning teacher uses middle-layer attenuation followed by upper-layer restoration. The middle/upper windows are 12–23/24–31 for Qwen3.5-4B, 12–34/35–41 for Gemma4-E4B, and 15–27/28–39 for Qwen3.5-35B-A3B, using zero-based indices. Coefficients are constant within each window and fixed before student training. The culture teacher

applies the residual projection in the middle-layer window. Table 2 reports the language-specific coefficients.

Table 2: Teacher intervention coefficients. Reasoning entries give the signed middle/upper coefficients (s_1, s_2); culture entries give the middle-layer coefficient $\alpha_{C,\ell}$.

	Reasoning teacher			Culture teacher		
Language	4B	Gemma	35B	4B	Gemma	35B
bn	$(0.15, 0)$	$(0.05, 0)$	$(0.2, 0)$	0	0.025	0
de	$(0.2, -0.15)$	$(0, -0.05)$	$(0, -0.2)$	0.05	0	0.1
es	$(0.15, -0.2)$	$(0, -0.05)$	$(0.15, -0.25)$	0.025	0	0.05
fr	$(0.2, -0.05)$	$(0, 0)$	$(0.1, -0.2)$	0.025	0	0
ja	$(0.05, 0)$	$(0, -0.05)$	$(0.15, -0.1)$	0.125	0	0
ru	$(0.15, -0.2)$	$(0, -0.1)$	$(0.15, -0.1)$	0	0	0
te	$(0, 0)$	$(0, 0)$	$(0.1, -0.15)$	0.05	0	0.15
zh	$(0.15, -0.05)$	$(0, -0.05)$	$(0, 0)$	0.025	0.025	0.15

E Evaluation Settings

Benchmarks and language coverage. MGSM, PolyMath, mAceReason-Math, and INCLUDE cover all eight languages, with 800, 1,200, 400, and 750 examples per method, respectively. CulturALL covers bn/es/ja/ru/zh with 50 examples per language, and UniMoral covers es/ru/zh with 100 examples per language for each of its four tasks. The same benchmark inputs are used across methods.

Metrics. We report language-macro task scores, using PolyMath's difficulty weighting within each language. CulturALL answers are evaluated with DeepSeek Flash in non-thinking mode. UniMoral averages weighted F1 for AP, MTC, and FAA with BERTScore for CG. R-Avg. and C-Avg. are arithmetic means of the three reasoning and three culture-sensitive benchmarks. Language fidelity measures final-response agreement with the requested language over MGSM, PolyMath, and INCLUDE, averaged equally across the eight languages.

F Language-Level Results

Figures 3, 6, and 7 report reasoning, culture-sensitive performance, and language fidelity by language. Each language's reasoning average uses all three mathematical benchmarks; its culture average uses the benchmarks available in that language.

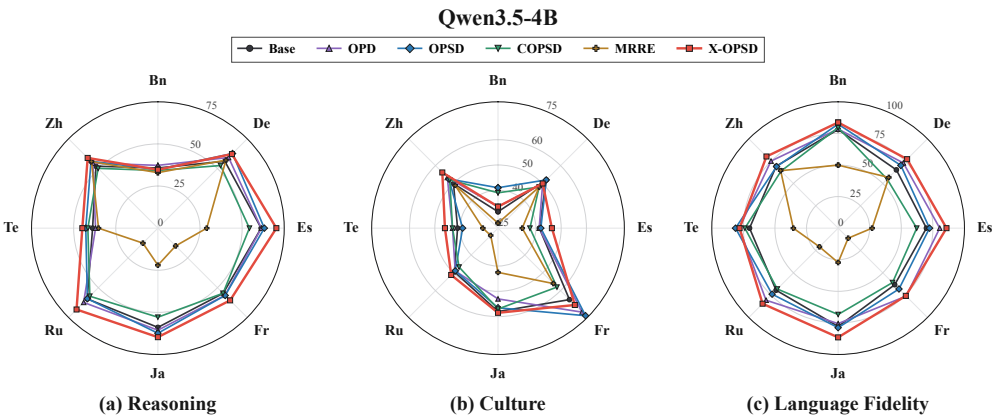


Figure 6: Qwen3.5-4B language-level results: reasoning, culture-sensitive performance, and language fidelity. Each panel uses a shared radial scale across languages and methods.

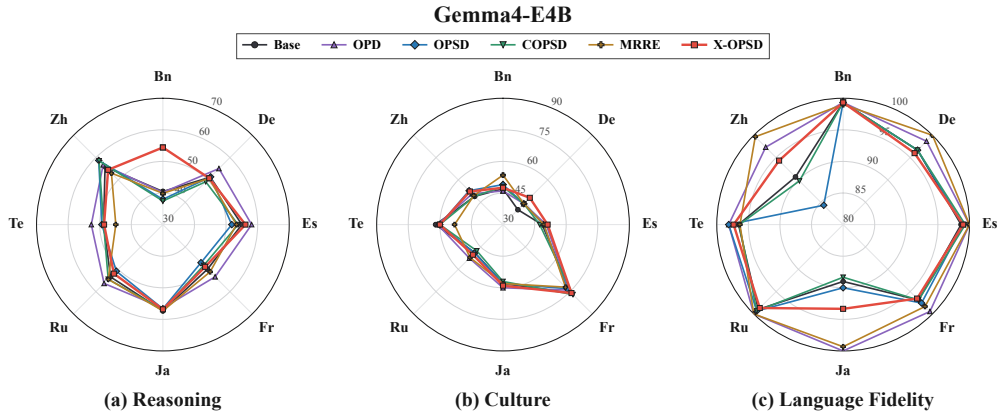


Figure 7: Gemma4-E4B language-level results, with the same three metrics and aggregation as Figure 6.

G Representation Analysis

Protocol. Figure 4(a,b) analyzes Qwen3.5-4B on 400 mAceReason-Math prompts and 750 held-out INCLUDE prompts. Each target-language question is paired with an English counterpart. We extract the last input token's state after each decoder block and project it into the corresponding fixed rank-7 task subspace. X-OPSD and COPSD are evaluated without steering; MRRE uses its inference-time intervention.

Let $d_{m,\ell}^{(k,S)}$ denote the mean paired-question Euclidean distance from method m to the English Base reference at layer k, in task subspace S and language ℓ. We report

$$D_m^{(k,S)} = \frac{100}{|\mathcal{L}|} \sum_{\ell \in \mathcal{L}} \frac{d_{m,\ell}^{(k,S)}}{d_{\mathrm{Base},\ell}^{(k,S)}}. \tag{8}$$

The target-language Base is therefore 100. The plotted values average consecutive groups of four layers after language aggregation.

H Component Ablations

We compare four additional variants of X-OPSD on Qwen3.5-4B. **Math-only** trains only on mathematical prompts with the reasoning teacher. **No protection** replaces the culture route's residual directions with the original mathematical directions while retaining its coefficients. **Reasoning only** applies the reasoning teacher to both tasks. **Residual-all** applies residual directions to both tasks while retaining each route's coefficients. The middle two correspond to the two variants shown in Figure 4(c).

Table 3: Component ablations on Qwen3.5-4B (%). R-Avg./C-Avg. average the three benchmarks in each group. F is paired CulturALL forgetting (lower is better).

Method	R-Avg.	C-Avg.	MGSM	PolyMath	mAce	INCLUDE	CulturALL	UniMoral	F
X-OPSD	58.08	46.95	65.63	46.11	62.50	65.63	29.20	46.03	20.00
Math-only	55.19	44.50	64.50	40.57	60.50	63.50	26.80	43.21	27.69
No protection	57.37	44.08	65.75	45.61	60.75	61.88	25.60	44.77	26.15
Reasoning only	55.94	46.17	63.50	44.57	59.75	63.38	28.00	47.13	26.15
Residual-all	53.86	42.65	64.88	42.96	53.75	63.13	23.20	41.63	36.92

I TEACHER-PREFERENCE ANALYSIS

Protocol. For each backbone, we use 10 MGSM questions per language and 128 completion positions per question. All models score identical mathematical prefixes. At each position, Top-10 overlap is the fraction of tokens shared between the two models' highest-ranked sets. Equation 7 compares the student's Base-adjusted preference for the reasoning teacher against the same-language culture teacher. Confidence intervals use paired bootstrap resampling at the question level.

Table 4: Reasoning-teacher preference by language. Δ_{route} is the Base-adjusted reasoning-versus-culture preference in percentage points; intervals are question-bootstrap 95% confidence intervals. † marks zero reasoning intervention.

Language	Δ_{route}	95% interval
Qwen3.5-4B		
bn	7.41	$[6.84, 7.98]$
de	5.13	$[4.40, 5.80]$
es	5.57	$[4.86, 6.25]$
fr	5.66	$[4.84, 6.52]$
ja	-1.25	$[-1.47, -1.02]$
ru	7.54	$[7.30, 7.75]$
te†	-3.13	$[-3.43, -2.81]$
zh	5.00	$[4.59, 5.36]$
Gemma4-E4B		
bn	1.34	$[0.94, 1.72]$
de	4.02	$[3.59, 4.48]$
es	3.79	$[3.53, 4.08]$
fr†	0.00	$[0.00, 0.00]$
ja	5.08	$[4.40, 5.70]$
ru	4.23	$[3.73, 4.74]$
te†	0.00	$[0.00, 0.00]$
zh	1.95	$[1.59, 2.38]$
Qwen3.5-35B-A3B		
bn	10.79	$[9.81, 11.90]$
de	0.13	$[-0.51, 0.78]$
es	5.68	$[5.04, 6.32]$
fr	9.27	$[8.86, 9.64]$
ja	8.55	$[7.85, 9.30]$
ru	10.74	$[9.87, 11.59]$
te	3.17	$[1.82, 4.85]$
zh†	-6.16	$[-6.73, -5.62]$

J MEASURING TRANSFER SPILLOVER

Paired forgetting. For item i, let $c_i^B, c_i^M \in \{0, 1\}$ denote correctness for Base and method M. We define

$$F(M \mid B) = 100 \, \frac{\sum_i c_i^B (1 - c_i^M)}{\sum_i c_i^B}. \tag{9}$$

We evaluate the same 250 CulturALL items per backbone, matching predictions by question and language. We also count newly correct answers to distinguish retention from the net balance of losses and gains.

Table 5: Paired CulturALL retention. N_B is the number of Base-correct items; Lost/New count correct-to-incorrect and incorrect-to-correct transitions. F is the forgetting rate.

Backbone	Method	N_B	Lost	New	F (%) $\downarrow$
Qwen3.5-4B	COPSD	65	17	24	26.15
	X-OPSD	65	13	21	20.00
Gemma4-E4B	COPSD	71	14	13	19.72
	MRRE	71	12	23	16.90
	X-OPSD	71	7	13	9.86
Qwen3.5-35B-A3B	COPSD	81	37	13	45.68
	X-OPSD	81	23	26	28.40