DCRD: RETRIEVAL-GROUNDED CURRICULUM DISTILLATION FOR IMBALANCED MULTI-DOMAIN QUESTION ANSWERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Knowledge distillation has become an effective approach for transferring the capabilities of large language models into smaller and more efficient students. However, existing distillation methods typically assume a single teacher and balanced training distributions, which limits their applicability to realistic question-answering scenarios involving heterogeneous domains with severe data imbalance. We present DCRD, a multi-teacher distillation framework for imbalanced multi-domain question answering that treats each domain–teacher pair as a distinct learning course and dynamically schedules training effort based on the student's recent validation accuracy on each course. The framework combines retrieval-augmented knowledge grounding—enriching teacher supervision with external evidence from Wikipedia and PubMed—with a curriculum scheduler that adaptively adjusts course sampling and training intensity according to the student's performance. Experiments on MedQA-USMLE, OpenBookQA, and StrategyQA show that our approach consistently improves performance across Flan-T5 model scales, reaching 43.29% average accuracy on Flan-T5-Large and outperforming teacher ensemble and MoE-based distillation baselines without additional inference cost. These results demonstrate that curriculum-aware multi-teacher distillation provides an effective and scalable solution for knowledge-intensive multi-domain reasoning. Our code is available at https://github.com/linda-2-hh/DCRD.

1 INTRODUCTION

Large language models (LLMs) demonstrate strong performance across diverse question answering (QA) tasks, including medical, scientific, and commonsense reasoning. However, their high computational cost and inference latency limit real-world deployment. Knowledge distillation (KD) offers a practical solution by transferring capabilities from large teacher models into compact students. Existing KD methods largely assume a single teacher, a single domain, and a balanced distribution. These assumptions rarely hold in practical QA systems, which span heterogeneous domains with distinct knowledge characteristics and severe data imbalance.

This mismatch introduces two core challenges. First, heterogeneous knowledge transfer is difficult: a single teacher rarely excels across all domains, while naïvely aggregating multiple teachers via ensembles or Mixture-of-Experts (MoE) ignores domain specialization and introduces conflicting supervision. Second, cross-domain imbalance destabilizes optimization, as dominant datasets overwhelm training updates, causing under-learning in low-resource domains. Furthermore, standard KD overlooks the external knowledge supporting teacher reasoning, which is crucial for knowledge-intensive QA. While prior work explores retrieval-augmented KD (Kang et al., 2023; Zhang et al., 2023) and multi-teacher KD (Tian et al., 2025; Chen et al., 2024), they typically assume balanced, single-domain settings, failing to address the interaction between heterogeneous knowledge and severe data imbalance.

To address these limitations, we present DCRD, a multi-teacher distillation framework for imbalanced multi-domain QA. DCRD treats each domain–teacher pair as an independent learning "course" and dynamically allocates training effort based on the student's recent performance. It integrates two complementary components. First, we extend retrieval-augmented distillation to the multi-teacher

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.

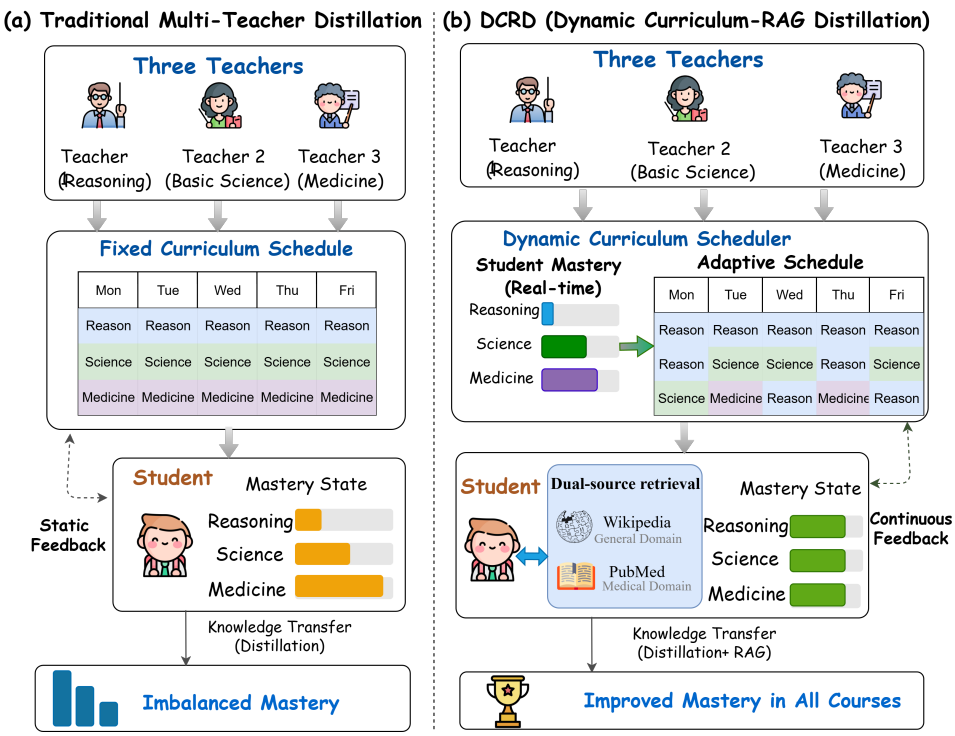


Figure 1: Conceptual comparison between conventional multi-teacher distillation and the proposed DCRD framework.

imbalanced scenario. Unlike prior works such as KARD and ReAugKD focusing on single teachers, we integrate external knowledge retrieved from Wikipedia and PubMed to ground supervision, training the student to reason with evidence rather than merely mimicking outputs. Second, a performance-driven curriculum scheduler adaptively adjusts course sampling, local epochs, and gradient accumulation to focus computation on under-performing domains, effectively mitigating severe data imbalance.

We evaluate DCRD on three heterogeneous QA benchmarks: MedQA-USMLE, OpenBookQA, and StrategyQA. Across Flan-T5 model scales, DCRD consistently outperforms standard distillation, teacher ensembles, and MoE routing, maintaining a single deployable student model without additional inference cost.

Our contributions are summarized as follows:

- We formulate multi-domain KD as a curriculum learning problem over domain-specific courses, treating each domain–teacher pair as an adaptive learning unit.
- We present a retrieval-augmented multi-teacher distillation framework that grounds teacher supervision with external evidence, enabling knowledge-aware student learning.
- We introduce a performance-driven curriculum scheduler that dynamically allocates training resources based on course-wise performance signals, mitigating cross-domain data imbalance.
- We demonstrate consistent improvements across heterogeneous QA benchmarks, showing that curriculum-aware and knowledge-grounded distillation provides an effective solution for scalable multi-domain reasoning.

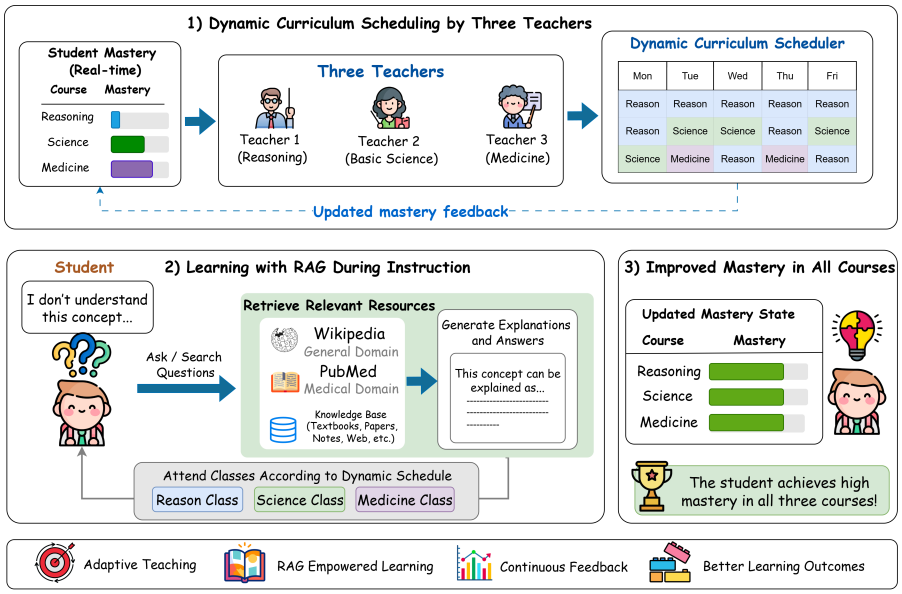


Figure 2: Overview of the proposed RAG-enhanced distillation framework. Teachers retrieve external knowledge from Wikipedia or PubMed to generate reasoning-aware supervision for student training.

2 Related Work

2.1 Knowledge Distillation for Language Models

Knowledge distillation (KD) compresses large language models into compact students. Early work transfers logits from a single teacher Hinton et al. (2015); Sanh et al. (2019), later extended to reasoning via chain-of-thought Wei et al. (2022a); Ho et al. (2023). Multi-teacher distillation has been extensively studied in vision You et al. (2017); Zhang et al. (2018) and NLP Gu et al. (2024). Traditional approaches typically aggregate teacher logits via weighting or voting, implicitly assuming homogeneous teacher competencies and balanced data. Recent works like TinyLLM Tian et al. (2025) and MAGDi Chen et al. (2024) transfer multi-teacher reasoning, yet they still operate under balanced settings. Unlike these mature paradigms that rely on prediction-level aggregation or instance-wise routing, DCRD performs parameter-level transfer through a shared student and explicitly addresses domain imbalance via dynamic curriculum scheduling.

2.2 Retrieval-Augmented Knowledge Distillation

Retrieval-augmented generation (RAG) Lewis et al. (2020) improves factual consistency by integrating external knowledge. Several mature studies have successfully extended retrieval to KD. KARD Kang et al. (2023) generates teacher rationales augmented with retrieved knowledge, while ReAugKD Zhang et al. (2023) combines a non-parametric memory with teacher soft labels. While effective, these established methods are primarily designed for single-teacher, single-domain scenarios. In contrast, our framework explicitly extends retrieval-augmented distillation to a multi-teacher imbalanced setting, where retrieved evidence serves as shared contextual grounding across heterogeneous domains and multiple teachers.

2.3 Multi-Teacher and Multi-Expert Routing

Beyond standard aggregation, Mixture-of-Experts (MoE) Shazeer et al. (2017) provides an alternative routing mechanism by dynamically assigning inputs to specialized experts. However, MoE routing does not inherently address data imbalance across domains and can suffer from catastrophic instability under severely skewed distributions, as shown in our experiments. Unlike MoE's hard or learned expert assignment, DCRD avoids instance-wise routing noise by isolating teachers at the course level, dynamically adjusting training effort based on course-wise performance rather than input features.

2.4 Curriculum Learning for Distillation

Curriculum learning Bengio et al. (2009) improves training by controlling sample order or difficulty Graves et al. (2017); Sachan & Xing (2016). In KD, strategies often focus on sample difficulty or mixing synthetic data Chen et al. (2026). Meanwhile, building unified LLMs often relies on multi-task training Sanh et al. (2021); Aribandi et al. (2021), which requires large-scale data and does not explicitly transfer teacher reasoning styles under imbalance. Our scheduler distinguishes itself by using course-wise validation accuracy—rather than heuristic difficulty scores—to drive sampling and training intensity, offering a simple and interpretable mechanism tailored for multi-domain distillation.

3 Proposed Method: The DCRD Framework

The proposed DCRD framework trains a compact student model using data drawn from multiple heterogeneous courses, where each course corresponds to a distinct domain–teacher pair (e.g., medical questions supervised by a medical-specialized teacher). The framework operates through three tightly coupled mechanisms: (1) retrieval-augmented distillation, which enriches each training instance with external factual evidence; (2) real-time learning state tracking, which monitors how well the student is progressing on each course; and (3) performance-driven curriculum scheduling, which adaptively allocates training resources—including sampling frequency, local training epochs, and gradient accumulation—based on the tracked states.

3.1 Retrieval-Augmented Distillation

For every question in a training batch, we first retrieve external knowledge to ground the distillation process. The retrieval source is selected according to the course domain: we query PubMed for medical reasoning tasks (MedQA-USMLE) and Wikipedia for scientific and commonsense reasoning tasks (OpenBookQA and StrategyQA). For each question, the retriever fetches the top three most relevant documents using dense passage retrieval, then truncates them to fit the model's input length.

These retrieved documents are combined with the original question and the teacher-generated explanation into a unified text prompt. Specifically, the student input is constructed by concatenating a task instruction prefix, the retrieved document texts, the question, and the teacher's reasoning chain. The student model is then trained to generate the teacher's final answer using standard sequence-to-sequence cross-entropy loss. By conditioning the student on both the retrieved evidence and the teacher's reasoning, this design encourages the model to learn verifiable, contextually grounded reasoning patterns rather than merely imitating surface-level outputs.

3.2 Tracking Learning Progress via Moving-Average Loss

To decide which course needs more training attention, we need a lightweight and real-time indicator of the student's performance on each domain. Computing exact validation accuracy during training would introduce significant overhead and delay scheduling decisions. Instead, we directly use the recent distillation loss as a proxy for learning progress.

For each course C_i, we maintain a moving average of the per-batch distillation losses observed over the most recent M training steps:

$$\bar{\ell}_i = \frac{1}{M} \sum_{t'=t-M+1}^{t} \mathcal{L}_{\text{KD},i}(t'), \tag{1}$$

where $\mathcal{L}_{\text{KD},i}(t')$ is the distillation loss on a batch sampled from course C_i at step t'. We set $M = 20$ in our experiments, which provides a smooth and stable estimate without being overly sensitive to batch-wise fluctuations. A lower $\bar{\ell}_i$ indicates that the student is aligning well with the teacher on that domain, while a higher value suggests that the student still struggles with the domain-specific reasoning patterns. This moving-average loss serves as the central signal that drives all subsequent scheduling decisions.

3.3 Performance-Driven Course Sampling

At the beginning of each global training iteration, the scheduler evaluates the moving-average loss for every course and determines which course to sample next. The core idea is simple: courses where the student's loss is high should be visited more frequently.

To implement this, we first define a target loss threshold λ that represents an acceptable level of proficiency. Then, for each course C_i, we compute a deficit score as the gap between its current moving-average loss and this threshold:

$$W_i = \max(0.01,\ \bar{\ell}_i - \lambda). \tag{2}$$

The clipping at 0.01 ensures that even well-learned courses retain a minimal chance of being sampled. Courses with larger deficits—meaning the student is further from the target—receive proportionally larger scores. We then convert these scores into a probability distribution by normalizing them across all courses:

$$P(C_i) = \frac{W_i}{\sum_{j=1}^{N} W_j}. \tag{3}$$

Finally, we sample the next course according to this distribution. This sampling strategy naturally shifts the student's training focus toward underperforming domains. Unlike static or uniform sampling, this mechanism adapts continuously as the student improves, gradually re-balancing the training effort across courses without requiring any manual adjustment of sampling ratios.

3.4 Adaptive Training Intensity for Selected Courses

Beyond adjusting the sampling frequency, our framework further modulates the training intensity for the selected course before the actual gradient update begins. Two aspects of training are dynamically controlled: the number of local epochs and the number of gradient accumulation steps.

Dynamic Local Epochs. Once a course is selected, instead of performing just a single pass over a mini-batch, we may repeat the update multiple times before returning to the global scheduler. We increase the number of local epochs for courses with higher moving-average losses. Starting from a base value of one epoch, we add extra epochs proportional to the deficit:

$$E_i = \max\left(1,\ \text{BASE_EPOCHS} + \lfloor(\bar{\ell}_i - \lambda) \cdot \Delta E\rfloor\right), \tag{4}$$

where $\Delta E = 5$ controls the scaling. This means a difficult course receives multiple consecutive gradient updates in a single visit—for example, if the loss exceeds the threshold by 0.2, the course may receive two local epochs instead of one. This allows the student to make more substantial progress on that domain before the scheduler re-evaluates the global sampling distribution.

Dynamic Gradient Accumulation. For courses that are inherently noisy or have limited training examples, we also adjust the effective batch size by modifying the number of gradient accumulation steps. When the moving-average loss is high, we increase the accumulation steps, which effectively simulates a larger batch size. This smooths out gradient estimates and stabilizes optimization on difficult or data-scarce domains. The accumulation step for course C_i is computed as:

$$\begin{aligned}\text{GradAcc}_i = \text{clamp}\big(&\text{MIN_ACC} + \lfloor(\bar{\ell}_i - \lambda) \cdot \Delta A\rfloor,\\ &\text{MIN_ACC},\ \text{MAX_ACC}\big),\end{aligned} \tag{5}$$

where we set $\text{MIN_ACC} = 4$ and $\text{MAX_ACC} = 12$, with $\Delta A = 8$ controlling the sensitivity. The clamp function keeps the value within the predefined range. In practice, this means that courses where the student struggles most effectively receive a larger effective batch size, providing more stable gradient signals, while well-learned courses are trained with smaller accumulations to maintain efficiency.

Model	Method	MedQA	OpenBookQA	StrategyQA	Average
77M	Few-shot Brown et al. (2020) (NeurIPS 2020)	27.73$^{-0.02}$	27.40$^{+0.92}$	49.20$^{-0.31}$	33.66$^{+1.60}$
	Few-shot + CoT Wei et al. (2022b) (NeurIPS 2022)	27.73$^{-0.02}$	26.40$^{-0.08}$	51.24$^{+1.73}$	34.03$^{+1.97}$
	KA. Few-shot + CoT	27.73$^{-0.02}$	**30.00**$^{+3.52}$	50.07$^{+0.56}$	**34.43**$^{+2.38}$
	Fine-tuning (LoRA) Hu et al. (2021) (ICLR 2021)	27.65$^{-0.10}$	28.00$^{+1.52}$	50.66$^{+1.15}$	34.15$^{+2.09}$
	MoE Shazeer et al. (2017) (ICLR 2017)	20.66$^{-7.09}$	18.80$^{-7.68}$	46.14$^{-3.37}$	27.40$^{-4.66}$
	self-KD Zhang et al. (2019) (arXiv)	20.90$^{-6.85}$	20.80$^{-5.68}$	51.67$^{+2.16}$	29.47$^{-2.59}$
	Ensemble Jiang et al. (2022) (ACL 2022)	20.82$^{-6.93}$	20.80$^{-5.68}$	48.47$^{-1.04}$	28.54$^{-3.52}$
	prompt-KD Gu et al. (2024) (ICLR 2024)	20.50$^{-7.25}$	21.40$^{-5.08}$	**51.82**$^{+2.31}$	29.43$^{-2.63}$
	BOLT Lu et al. (2025) (Appl. Sci. 2025)	27.73$^{-0.02}$	27.40$^{+0.92}$	49.20$^{-0.31}$	33.66$^{+1.60}$
	RTKD Son & Kim (2026) (KBS 2026)	27.73$^{-0.02}$	27.00$^{+0.52}$	51.38$^{+1.87}$	34.19$^{+2.13}$
	DCRD (ours)	**27.97***$^{+0.22}$	27.80$^{+1.32}$	48.47$^{-1.04}$	33.66$^{+1.60}$
250M	Few-shot Brown et al. (2020) (NeurIPS 2020)	27.73$^{+0.05}$	41.40$^{+2.75}$	38.14$^{-4.34}$	33.42$^{+0.29}$
	Few-shot + CoT Wei et al. (2022b) (NeurIPS 2022)	26.79$^{-0.89}$	20.80$^{-17.85}$	51.53$^{+9.05}$	32.48$^{-0.65}$
	KA. Few-shot + CoT	26.16$^{-1.52}$	**46.00**$^{+7.35}$	51.38$^{+8.90}$	37.24$^{+4.11}$
	Fine-tuning (LoRA) Hu et al. (2021) (ICLR 2021)	27.89$^{+0.21}$	41.20$^{+2.55}$	21.40$^{-21.08}$	28.78$^{-4.35}$
	MoE Shazeer et al. (2017) (ICLR 2017)	27.18$^{-0.50}$	26.40$^{-12.25}$	3.78$^{-38.70}$	20.49$^{-12.64}$
	self-KD Zhang et al. (2019) (arXiv)	20.97$^{-6.71}$	32.80$^{-5.85}$	54.59$^{+12.11}$	32.76$^{-0.37}$
	Ensemble Jiang et al. (2022) (ACL 2022)	**29.30**$^{+1.62}$	43.40$^{+4.75}$	49.05$^{+6.57}$	37.68$^{+4.55}$
	prompt-KD Gu et al. (2024) (ICLR 2024)	22.94$^{-4.74}$	38.60$^{-0.05}$	**55.31**$^{+12.83}$	35.16$^{+2.03}$
	BOLT Lu et al. (2025) (Appl. Sci. 2025)	27.73$^{+0.05}$	41.40$^{+2.75}$	38.14$^{-4.34}$	33.42$^{+0.29}$
	RTKD Son & Kim (2026) (KBS 2026)	27.73$^{+0.05}$	37.60$^{-1.05}$	46.58$^{+4.10}$	35.00$^{+1.87}$
	DCRD (ours)	27.49$^{-0.19}$	45.60*$^{+6.95}$	51.97$^{+9.49}$	**38.01***$^{+4.88}$
780M	Few-shot Brown et al. (2020) (NeurIPS 2020)	27.73$^{-0.49}$	57.60$^{+5.65}$	56.91$^{+5.66}$	41.95$^{+2.80}$
	Few-shot + CoT Wei et al. (2022b) (NeurIPS 2022)	27.73$^{-0.49}$	26.40$^{-25.55}$	51.38$^{+0.13}$	34.06$^{-5.09}$
	KA. Few-shot + CoT	30.56$^{+2.34}$	54.00$^{+2.05}$	48.18$^{-3.07}$	40.24$^{+1.10}$
	Fine-tuning (LoRA) Hu et al. (2021) (ICLR 2021)	29.85$^{+1.63}$	54.80$^{+2.85}$	53.42$^{+2.17}$	41.50$^{+2.35}$
	MoE Shazeer et al. (2017) (ICLR 2017)	26.79$^{-1.43}$	26.60$^{-25.35}$	51.53$^{+0.28}$	33.66$^{-5.49}$
	self-KD Zhang et al. (2019) (arXiv)	25.45$^{-2.77}$	50.80$^{-1.15}$	**64.63**$^{+13.38}$	41.54$^{+2.40}$
	Ensemble Jiang et al. (2022) (ACL 2022)	29.30$^{+1.08}$	54.60$^{+2.65}$	49.78$^{-1.47}$	40.16$^{+1.02}$
	prompt-KD Gu et al. (2024) (ICLR 2024)	28.12$^{-0.10}$	47.20$^{-4.75}$	52.69$^{+1.44}$	38.86$^{-0.29}$
	BOLT Lu et al. (2025) (Appl. Sci. 2025)	27.73$^{-0.49}$	57.60$^{+5.65}$	56.91$^{+5.66}$	41.95$^{+2.80}$
	RTKD Son & Kim (2026) (KBS 2026)	27.73$^{-0.49}$	41.40$^{-10.55}$	38.14$^{-13.11}$	33.42$^{-5.73}$
	DCRD (ours)	**31.26***$^{+3.04}$	**60.00***$^{+8.05}$	53.28$^{+2.03}$	**43.25***$^{+4.11}$

Table 1: Dataset sample counts: MedQA-USMLE=1273, OpenBookQA=500, StrategyQA=687, total samples=2460. Overall and per-dataset accuracy (%) of Flan-T5 variants. Averages are weighted by dataset size, calculated as total correct samples / total samples (2460). Superscripts denote deviation from the column mean. **Bold** and underline indicate best and second best. *: DCRD significantly outperforms the best baseline ($p < 0.05$).

Through this combination of performance-driven sampling, dynamic local epochs, and adaptive gradient accumulation, the DCRD framework provides a comprehensive and fine-grained mechanism for balancing multi-domain distillation. It ensures that the student model receives the right amount of training on the right domains at the right intensity, effectively mitigating the severe data imbalance that plagues real-world question-answering systems.

4 Experiments

4.1 Experimental Setup

Datasets and models. We evaluate on three question-answering benchmarks: MedQA-USMLE Jin et al. (2021), OpenBookQA Mihaylov et al. (2018), and StrategyQA Geva et al. (2021). They cover medical, scientific, and multi-step commonsense reasoning. The student is Flan-T5 Raffel et al. (2023) at three scales: Small (77M), Base (250M), and Large (780M). The domain teachers are GPT Floridi & Chiriatti (2020) for MedQA-USMLE, GLM-4.5-air Zeng et al. (2025) for OpenBookQA, and DeepSeek-R1-0528 Guo et al. (2025) for StrategyQA. Retrieval uses PubMed for MedQA-USMLE and Wikipedia for the other two datasets.

Table 2: Zero-shot accuracy (%) on five unseen MMLU subjects using Flan-T5-Large. The average values are retained as reported in the source results.

Method	HS Chem.	HS Phys.	Psych.	Anat.	Clinical	Average
0-shot	25.25	26.00	52.08	41.22	46.91	–
CoT	27.72	26.00	52.67	40.46	43.21	42.08
LoRA	27.72	24.67	54.46	41.22	46.91	43.46
self-KD	28.22	29.33	48.71	38.93	45.68	41.27
Ensemble	28.71	28.67	52.08	41.98	45.68	43.05
MoE	28.22	28.67	28.51	23.66	26.34	39.34
DCRD	26.73	25.33	**54.46**	40.46	**47.74**	**43.54**

Table 3: Component ablations across Flan-T5 scales (accuracy, %). "w/o mt" denotes the reported single-teacher variant; "w/o rag" removes retrieval; "w/o dy" uses uniform sampling.

Method	Small (77M)	Base (250M)	Large (780M)	Average
w/o mt	30.12	26.46	24.31	26.96
w/o rag	**34.15**	36.50	39.87	36.84
w/o dy	32.44	37.24	41.06	36.91
DCRD	33.66	**38.00**	**43.29**	**38.32**

Evaluation and baselines. We report exact-match accuracy on each dataset and the overall accuracy weighted by test-set size. The moving-average loss is used to schedule training courses, not to score test examples. The comparisons cover prompting (few-shot, few-shot with chain-of-thought, and knowledge-augmented few-shot), LoRA fine-tuning Hu et al. (2021), self-KD Zhang et al. (2019), prompt-KD Gu et al. (2024), BOLT Lu et al. (2025), RTKD Son & Kim (2026), teacher ensembling Jiang et al. (2022), and sparse-gated MoE routing Shazeer et al. (2017). Trainable student baselines use the Flan-T5 backbone and the same multi-domain data; prompting and teacher ensembling provide inference-time reference points.

Implementation. Training uses AdamW on NVIDIA A40 GPUs. The software environment comprises CUDA 12.4, PyTorch 2.6.0, Transformers 4.49.0, and Python 3.12.9. Experiments are repeated with three random seeds. Training and evaluation prompts are provided in the supplementary material.

4.2 Main Results and Generalization

Table 1 gives the full in-domain comparison. DCRD obtains the highest reported overall accuracy for the Base and Large students. The gain is less uniform for Small: its 33.66% average does not exceed every baseline in that block. We therefore focus the remaining analysis on which components account for the gains at larger student capacity, rather than treating the result as uniform across all scales and datasets. We also evaluate Flan-T5-Large without additional training on five unseen MMLU subjects. Table 2 reports the subject-level results. DCRD has the highest reported overall score (43.54%) and the best clinical score, while other methods remain competitive on individual subjects. This suggests some transfer beyond the training domains, although the evidence is mixed across subjects.

4.3 Component Analysis

Table 3 isolates the reported effects of teacher configuration, retrieval, and dynamic scheduling. On the Large student, removing retrieval reduces average accuracy to 39.87%, and replacing dynamic scheduling with uniform sampling reduces it to 41.06%. Retrieval has a different effect on Small, where the no-retrieval variant reaches 34.15%, above the full model's 33.66%. Thus its benefit should not be described as uniform across model sizes. To examine the scheduler directly, we compare dynamic course selection with static curriculum ordering and random sampling under the same curriculum-comparison setting. In Table 4, dynamic scheduling improves the reported average from 40.53% (static) and 40.40% (random) to 41.87%. The largest subject-level difference is on StrategyQA; random sampling remains stronger on MedQA-USMLE.

Table 4: Curriculum comparison on Flan-T5-Large (accuracy, %).

Curriculum strategy	MedQA	OpenBookQA	StrategyQA	Average
Static curriculum	28.67	53.60	52.98	40.53
Random sampling	**30.16**	54.60	49.05	40.40
DCRD (dynamic)	29.69	**54.80**	**55.02**	**41.87**

Table 5: Retrieval settings on Flan-T5-Large (accuracy, %).

Setting	MedQA	OpenBookQA	StrategyQA	Average
No RAG	28.12	47.20	52.69	38.90
Teacher-RAG only	30.79	53.40	50.22	40.81
Teacher+student RAG	30.40	49.20	48.33	39.23
DCRD (full)	**31.26**	**60.00**	**53.28**	**43.25**

4.4 Retrieval and Data Imbalance

We retrieve the top three documents using the question as the query. Table 5 separates teacher-side retrieval, student-side retrieval, and the full system. Teacher-side retrieval improves the reported average over no retrieval (40.81% versus 38.90%). Simply adding retrieval to both teacher and student does not improve on teacher-side retrieval in this comparison; the full DCRD setting attains the highest reported average. These results support evaluating retrieval together with the remaining training components rather than attributing the full gain to retrieval alone. We next vary the training-data ratio across StrategyQA, MedQA-USMLE, and OpenBookQA from 1:1:1 to 1:8:16, while keeping the Flan-T5-Large architecture and training setup fixed. Figure 3 compares DCRD with random sampling, loss reweighting, and focal loss. The plotted results show that DCRD degrades less as the imposed cross-domain imbalance increases.

4.5 Teacher Aggregation and Conflicting Supervision

DCRD transfers supervision into a shared student rather than combining teacher predictions at inference time. Table 8 summarizes the mechanisms, and Table 6 compares their reported accuracy. On the Small and Large students shown here, DCRD has a higher average than teacher ensembling and MoE routing. The comparison supports the value of the shared-student training procedure, but does not by itself isolate which individual component produces the difference. We also replace 50% of explanations with explanations from incorrect-domain teachers. Table 7 shows that DCRD's reported average decreases from 43.29% to 42.08%. Under this intervention it is nearly tied with MoE (42.07%); the effect varies across datasets. This experiment therefore provides a stress test of conflicting supervision rather than evidence that one method is uniformly unaffected by it.

4.6 Training Cost and Expert Evaluation

Table 9 reports training time and total FLOPs. DCRD has the lowest reported FLOP count among these methods, although its wall-clock training time is longer than that of several baselines. Its deployed student does not require a teacher ensemble or MoE routing module. In a separate blind evaluation by three domain experts, DCRD receives the highest reported correctness and reasoning preference percentages (Table 10).

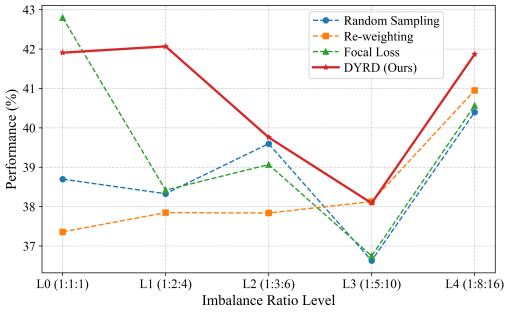


Figure 3: Accuracy under increasing cross-domain data imbalance. All methods use the same student architecture and training setup.

Table 6: Teacher ensemble, MoE routing, and DCRD on two Flan-T5 scales (accuracy, %).

Method	MedQA	OpenBookQA	StrategyQA	Average
		Flan-T5-Small		
Teacher ensemble	20.82	20.80	48.47	28.54
MoE	20.66	18.80	46.14	27.40
DCRD	**27.97**	**27.80**	**48.47**	**33.66**
		Flan-T5-Large		
Teacher ensemble	29.30	54.60	49.78	40.16
MoE	26.79	26.60	51.53	33.66
DCRD	**31.26**	**60.00**	**53.28**	**43.29**

Table 7: Accuracy (%) under clean and conflicting teacher explanations. The conflict condition replaces 50% of explanations with those from incorrect-domain teachers.

Setting	Method	MedQA	OpenBookQA	StrategyQA	Average
Clean	Teacher ensemble	29.30	54.60	49.78	40.16
	MoE	26.79	26.60	51.53	33.66
	DCRD	**31.26**	**60.00**	**53.28**	**43.29**
Conflict	Teacher ensemble	31.97	52.20	50.07	41.14
	MoE	**32.05**	**55.60**	50.80	42.07
	DCRD	29.46	55.20	**55.90**	**42.08**

5 Conclusion

We introduced DCRD, a unified framework for knowledge-aware adaptive distillation in multi-domain QA. By integrating retrieval-augmented grounding with a loss-driven curriculum scheduler, DCRD explicitly addresses heterogeneous domains, data imbalance, and inconsistent teacher supervision. Extensive experiments show DCRD consistently outperforms static distillation, ensembles, and MoE routing—maintaining a single deployable student without extra inference cost—suggesting that realistic QA distillation must be both knowledge-grounded and curriculum-aware. Despite these gains, dynamic online retrieval introduces training overhead and potential noise from irrelevant documents. Future work will explore efficient local vector databases, cached retrieval, and structured causal reasoning to reduce computational cost and retrieval noise.

Method	**Multi-Tch**	**Logit Avg**	**Curr**	**Rout**
Teacher Ensemble	✓	✓	×	×
MoE	✓	×	×	✓
DCRD(Ours)	✓	×	✓	✓

Table 8: Conceptual comparison of multi-teacher strategies. "Curr" indicates curriculum scheduling; "Rout" indicates explicit routing.

Table 9: Reported training cost.

Method	Training time	Total FLOPs
LoRA ($r = 8$)	14 min	11 EFLOPs
BOLT	∼1 h	7 EFLOPs
Ensemble	∼1.5 h	14 EFLOPs
self-KD	∼57 min	93 EFLOPs
prompt-KD	∼2 h	194 EFLOPs
MoE	5 h	116 EFLOPs
DCRD	∼1.6 h	**3.1 EFLOPs**

Table 10: Blind expert evaluation (%).

Method	Correctness win	Reasoning preference
Ensemble	63.3	58.9
MoE	60.0	55.6
DCRD	**83.3**	**86.7**

AI USE STATEMENT

We used generative AI tools solely to improve the readability, grammar, and language of the manuscript. No generative AI tools were used for research ideation, methodology or experiment design, implementation, data analysis, or interpretation. All AI-assisted text was reviewed and verified by the authors, who take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

We provide the implementation, experimental configurations, and supporting materials necessary to reproduce our results. The source code will be publicly available at `https://github.com/linda-2-hh/DCRD` upon publication. Additional implementation details, experimental settings, and supplementary results are provided in the appendix and supplementary materials.

REFERENCES

Vamsi Aribandi, Yi Tay, Tal Schuster, Jinfeng Rao, Huaixiu Steven Zheng, Sanket Vaibhav Mehta, Honglei Zhuang, Vinh Q Tran, Dara Bahri, Jianmo Ni, et al. Ext5: Towards extreme multi-task scaling for transfer learning. *arXiv preprint arXiv:2111.10952*, 2021.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1877–1901, 2020.

Dongjie Chen, Kartik Patwari, Zhengfeng Lai, Xiaoguang Zhu, Sen-ching Cheung, and Chen-Nee Chuah. Empowering source-free domain adaptation via mllm-guided reliability-based curriculum learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4262–4272, 2026.

Justin Chih-Yao Chen, Swarnadeep Saha, Elias Stengel-Eskin, and Mohit Bansal. Magdi: Structured distillation of multi-agent interaction graphs improves reasoning in smaller language models. *arXiv preprint arXiv:2402.01620*, 2024.

Luciano Floridi and Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. *Minds and machines*, 30(4):681–694, 2020.

Mor Geva, Daniel Khashabi, Elad Segal, Tushar Khot, Dan Roth, and Jonathan Berant. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies. *Transactions of the Association for Computational Linguistics*, 9:346–361, 2021.

Alex Graves, Marc G Bellemare, Jacob Menick, Remi Munos, and Koray Kavukcuoglu. Automated curriculum learning for neural networks. In *international conference on machine learning*, pp. 1311–1320. Pmlr, 2017.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation 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. 32694–32717, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/8ac015d409635f196f9e3e9dcfb9a94e-Paper-Conference.pdf.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Namgyu Ho, Laura Schmid, and Se-Young Yun. Large language models are reasoning teachers. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pp. 14852–14882, 2023.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Lu Wang, Weizhu Wang, and Bin Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Machine Learning*, pp. 261–279. PMLR, 2021.

Zhengbao Jiang, Simran Arora, Aditi Grover, Claire Zhang, and Diyi Yang. Improving language model fine-tuning via ensemble robustness. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 102–109, 2022.

Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14):6421, 2021.

Minki Kang, Seanie Lee, Jinheon Baek, Kenji Kawaguchi, and Sung Ju Hwang. Knowledge-augmented reasoning distillation for small language models in knowledge-intensive tasks. *Advances in Neural Information Processing Systems*, 36:48573–48602, 2023.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33: 9459–9474, 2020.

Runze Lu, Zhaoyu Fan, Guanjie Wang, and Qingjiang Shi. Bolt: Building open-source llms for your target domain via automated hierarchical knowledge distillation. *Applied Sciences*, 15(21):11393, 2025.

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 *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2381–2391, 2018.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2023. URL https://arxiv.org/abs/1910.10683.

Mrinmaya Sachan and Eric Xing. Easy questions first? a case study on curriculum learning for question answering. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 453–463, 2016.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, et al. Multitask prompted training enables zero-shot task generalization. *arXiv preprint arXiv:2110.08207*, 2021.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Geonyeong Son and Misuk Kim. Rtkd: Responsive teacher knowledge distillation with heterogeneous students. *Knowledge-Based Systems*, pp. 116101, 2026.

Yijun Tian, Yikun Han, Xiusi Chen, Wei Wang, and Nitesh V Chawla. Beyond answers: Transferring reasoning capabilities to smaller llms using multi-teacher knowledge distillation. In *Proceedings of the eighteenth acm international conference on web search and data mining*, pp. 251–260, 2025.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022a.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022b.

Shan You, Chang Xu, Chao Xu, and Dacheng Tao. Learning from multiple teacher networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1285–1294, 2017.

Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao Zeng, Jiajie Zhang, et al. Glm-4.5: Agentic, reasoning, and coding (arc) foundation models. *arXiv preprint arXiv:2508.06471*, 2025.

Jianyi Zhang, Aashiq Muhamed, Aditya Anantharaman, Guoyin Wang, Changyou Chen, Kai Zhong, Qingjun Cui, Yi Xu, Belinda Zeng, Trishul Chilimbi, et al. Reaugkd: Retrieval-augmented knowledge distillation for pre-trained language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 1128–1136, 2023.

Linfeng Zhang, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. Be your own teacher: Improve the performance of convolutional neural networks via self distillation. *arXiv preprint arXiv:1905.08094*, 2019.

Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu. Deep mutual learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4320–4328, 2018.

A Additional Analysis

We provide additional analyses to further illustrate the behavior and design choices of the proposed DCRD framework.

A.1 Comparison of overall accuracy (% ± std) across different model sizes

Our proposed DCRD method consistently outperforms all baseline methods on the Base (250M) and Large (780M) model variants, and achieves the highest average accuracy (38.32%) across the three benchmarks (MedQA-USMLE, OpenbookQA, StrategyQA). Notably, while some methods (e.g., Few-shot, BOLT) perform well on the Large model, they suffer from lower accuracy on the Small or Base models. In contrast, DCRD demonstrates balanced improvements across all model sizes, with particularly strong gains on the Base model (38.00% vs. the next best 37.68% from ensemble). The error bars (standard deviations over three runs) indicate that DCRD also yields stable performance, as evidenced by its small variance.

A.2 Dynamic Curriculum Scheduling Visualization.

Figure 5 visualizes the evolution of curriculum scheduling throughout training. The line chart depicts the changing sampling frequency of each course across training stages, while the heatmap illustrates the corresponding difficulty distribution inferred from pseudo-accuracy. Early training stages emphasize data-rich or easier courses to establish a stable knowledge foundation. As training progresses, DCRD gradually reallocates focus toward weaker or underperforming courses, demonstrating an explicit and interpretable weakness-driven learning trajectory. This visualization provides direct empirical evidence that DCRD dynamically adapts its curriculum rather than following a static or heuristic schedule.

A.3 Model and Data Scaling Analysis.

We further investigate the scalability of DCRD with respect to both model size and training data volume. As shown in Figure 6, DCRD consistently outperforms the fine-tuning baseline across different data budgets and model scales. Notably, DCRD achieves competitive or superior performance compared to larger models under few-shot settings, highlighting its efficiency in leveraging limited data. Due to space constraints, detailed scaling results are deferred to the Appendix. Additional results on the effect of training data size are provided in Appendix B.4.

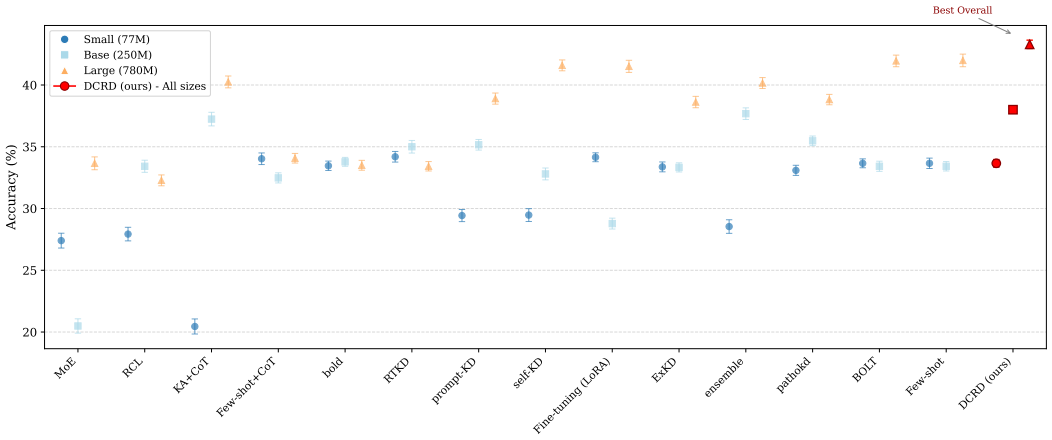


Figure 4: Comparison of overall accuracy (% ± std) across different model sizes (Small: 77M, Base: 250M, Large: 780M) on the combined set of MedQA-USMLE, OpenbookQA and StrategyQA using Flan-T5. Each method is shown with three markers corresponding to model sizes, with horizontal offsets for clarity. Our DCRD (highlighted in red) achieves the best average accuracy (38.32%) and leads on both Base and Large models. Error bars represent standard deviation over three independent runs.

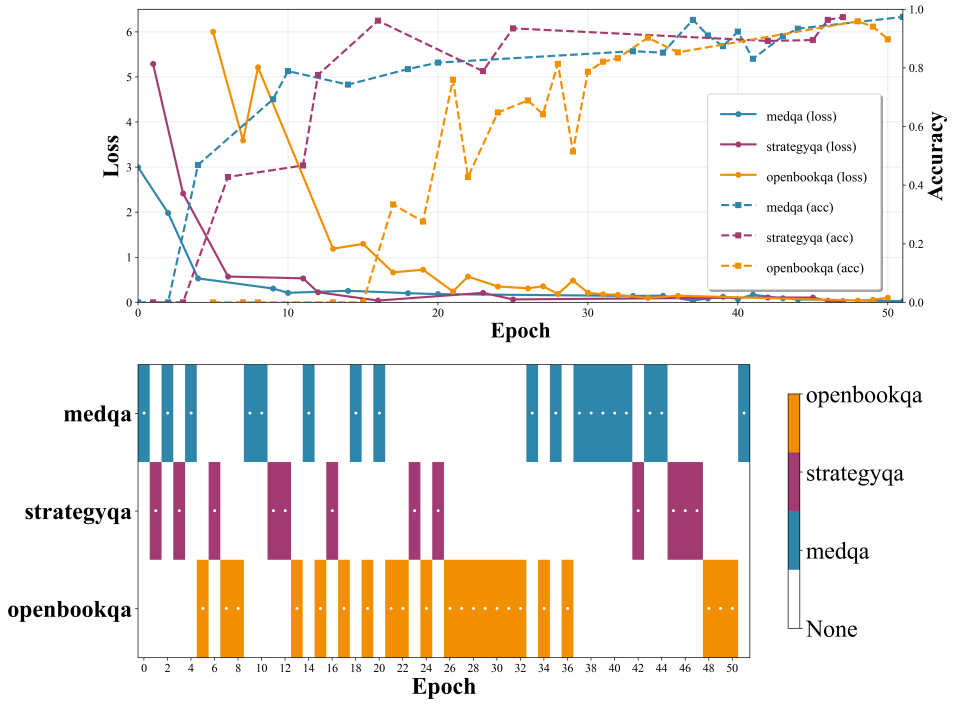


Figure 5: Line Chart and Heatmap of Dynamic Curriculum Scheduling: Sample Difficulty Distribution Across Stages

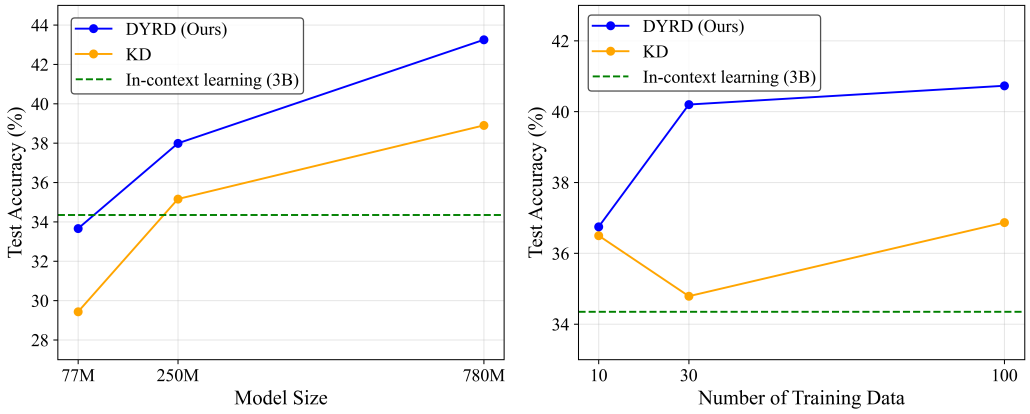


Figure 6: (a)Efficiency on training data and (b) model size. On MedQA-USMLE, OpenbookQA, and StrategyQA, we compare DCRD(ours) against the fine-tuning baseline by varying either the number of training data with Flan-T5 Large or the number of parameters, including the few-shot in-context learning performance of Flan-T5 XL(3B).

Table 11: Performance vs. Number of Retrieved Documents (K) on t5-large

Dataset	K=1	K=2	K=3	K=4
MedQA	28.04	30.48	31.26	30.48
OpenbookQA	55.00	54.80	60.00	55.40
StrategyQA	51.38	52.84	53.42	53.71
Overall	40.04	41.67	**43.29**	42.03

Note: The experiment evaluates the impact of the number of retrieved documents (default K=3) on model performance (t5-large backbone), with values representing accuracy (in percentage).

A.4 Effect of Retrieved Document Number (K).

We analyze the sensitivity of DCRD to the number of retrieved documents used during RAG-enhanced distillation. Table 11 reports performance on the T5-large backbone under different values of K. Overall performance improves as K increases from 1 to 3, indicating that additional retrieved evidence provides complementary factual grounding. However, further increasing K to 4 leads to marginal degradation, suggesting that excessive retrieval may introduce noise or distract the student model. Based on this trade-off, we adopt $K = 3$ as the default setting in all experiments.

A.5 Impact of Cross-Domain Imbalance Severity.

Table 12 reports the performance of different imbalance handling strategies under progressively increasing cross-domain skew. By artificially controlling the sampling ratios across domains (from L0 to L4), this study isolates the effect of imbalance severity while keeping the model architecture and training budget fixed.

Table 12: Performance under Different Cross-Domain Imbalance Levels (Flan-T5-Large)

Method	L0 (1:1:1)	L1 (1:2:4)	L2 (1:3:6)	L3 (1:5:10)	L4 (1:8:16)
Random Sampling	38.70	38.33	39.60	36.63	40.40
Loss Re-weighting	37.36	37.85	37.84	38.13	40.95
Focal Loss	**42.80**	38.42	39.07	36.75	40.57
DCRD (ours)	41.91	**42.07**	**39.76**	**38.09**	**41.87**

Method	Med	Openbook	Strategy	Avg.
		Flan-T5-Small		
Teacher Ens.	20.82	20.80	48.47	28.54
MoE Routing	20.66	18.80	46.14	27.40
DCRD	**27.97**	**27.80**	**48.47**	**33.66**
		Flan-T5-Base		
Teacher Ens.	29.30	43.40	49.05	37.68
MoE Routing	27.18	26.40	3.78	20.49
DCRD	**27.49**	**45.60**	**51.97**	**38.00**
		Flan-T5-Large		
Teacher Ens.	29.30	54.60	49.78	40.16
MoE Routing	26.79	26.60	51.53	33.66
DCRD	**31.26**	**60.00**	**53.28**	**43.29**

Table 13: Comparison with Teacher Ensemble and MoE-style Routing.

Table 14: Supplementary Ablation Results of DCRD across Model Scales and Datasets (%).

	MedQA-USMLE			OpenbookQA			StrategyQA			Overall		
Method	Small	Base	Large	Small	Base	Large	Small	Base	Large	Small	Base	Large
DCRD w/o MT	21.29	17.36	28.44	23.40	18.60	22.60	51.38	49.05	17.90	30.12	26.46	24.31
DCRD w/o Dynamic Curriculum	26.55	27.81	30.71	27.60	39.80	53.60	46.87	52.69	51.09	32.44	37.24	41.06
DCRD w/o RAG	27.26	26.39	30.01	27.60	43.40	52.20	51.67	50.22	49.20	34.15	36.50	39.87

As the imbalance becomes more extreme, conventional methods such as loss re-weighting and focal loss exhibit unstable or diminishing gains. In contrast, DCRD maintains consistently strong performance across all imbalance levels and achieves the best results in most settings. This trend indicates that DCRD effectively mitigates cross-domain imbalance through curriculum-aware training dynamics rather than static loss adjustments, demonstrating its robustness under severe distribution shifts.

A.6 Multi-Teacher Distillation Is Not Ensemble nor MoE

Table 13 compares DCRD with two representative multi-teacher paradigms: teacher ensemble and MoE-style routing. Although both baselines introduce multiple teachers, their improvements are limited under cross-domain imbalance. Teacher ensemble simply aggregates supervision from different teachers without explicitly considering domain difficulty, leading to insufficient adaptation to weak domains. MoE routing dynamically selects experts, but its performance becomes unstable when training data are highly imbalanced across domains.

In particular, the Flan-T5-Base student with MoE routing achieves only 3.78% accuracy on StrategyQA. We repeated this experiment multiple times and consistently observed similarly poor results. We hypothesize that the relatively small model capacity prevents the MoE router from learning stable expert allocation under severely imbalanced multi-domain supervision. As a result, weak domains are easily neglected during training, causing catastrophic performance collapse.

In contrast, DCRD explicitly models domain weakness through adaptive curriculum scheduling rather than relying on implicit expert routing. By dynamically reallocating training frequency, local epochs, and optimization intensity toward underperforming domains, DCRD maintains stable learning behavior across all model scales and achieves consistently stronger overall performance.

B Supplementary Intermediate Experimental Results

B.1 Per-Dataset Results of Flan-T5 Variants

This appendix provides detailed per-dataset results for the three primary benchmarks used in our main experiments: MedQA-USMLE, OpenbookQA, and StrategyQA. Tables 16, 17, and 18

Table 15: Supplementary Results under Different Cross-Domain Imbalance Levels (Flan-T5-Large, Accuracy %).

	L0 (1:1:1)			L1 (1:2:4)			L2 (1:3:6)			L3 (1:5:10)			L4 (1:8:16)		
Method	Med	OB	Str	Med	OB	Str	Med	OB	Str	Med	OB	Str	Med	OB	Str
Random Sampling	25.14	49.60	55.90	25.29	53.40	51.53	28.44	54.40	49.49	26.00	48.80	48.62	30.16	54.60	49.05
Loss Re-weighting	23.96	48.20	54.29	26.16	51.60	49.49	26.55	51.00	49.20	27.57	50.20	48.91	30.09	52.80	50.95
Focal Loss	30.32	55.40	56.77	25.77	51.20	52.55	27.26	54.20	49.93	25.53	48.60	48.91	30.01	55.40	49.34
DCRD (ours)	**28.59**	**53.80**	**57.93**	**28.59**	**55.80**	**57.06**	**27.10**	**54.00**	**52.84**	**26.08**	**52.60**	**49.78**	**29.69**	**54.80**	**55.02**

Table 16: MedQA-USMLE accuracy (%) of Flan-T5 variants.

Method	Flan-T5-Small	Flan-T5-Base	Flan-T5-Large	Avg.
Few-shot	27.73	27.73	27.73	27.73
Few-shot + CoT	27.73	26.79	27.73	27.42
KA. Few-shot + CoT	27.73	26.16	30.56	28.15
Fine-tuning (LoRA)	27.65	27.89	29.85	28.46
self-KD	20.90	20.97	25.45	22.44
prompt-KD	20.50	22.94	28.12	23.85
BOLT	27.73	27.73	27.73	27.73
RTKD	27.73	27.73	27.73	27.73
ours w/o dy	26.55	27.81	30.71	28.36
ours w/o RAG	27.26	26.39	30.01	27.89
ours	27.97	27.49	31.26	**28.91**

report the accuracy of different `Flan-T5` variants (Small, Base, and Large) under various training and distillation strategies.

Relation to Main Results. The overall accuracy reported in the main experimental section is computed as the macro-average over these three datasets. Specifically, for each method and model scale, the final score is obtained by averaging the corresponding accuracies on MEDQA-USMLE, OPENBOOKQA, and STRATEGYQA. We place these per-dataset breakdowns in the appendix to avoid redundancy while enabling a more fine-grained analysis of model behavior across domains.

Observations. Across all three benchmarks, our method consistently achieves the strongest or highly competitive performance, particularly on the `Flan-T5-Large` backbone. The ablated variants (*w/o dy* and *w/o RAG*) show clear performance degradation on at least one dataset, confirming that dynamic curriculum scheduling and retrieval-augmented distillation contribute complementary benefits. Moreover, the performance trends are largely consistent across model scales, indicating that the gains of our approach are not limited to a specific parameter regime.

Table 17: OpenbookQA accuracy (%) of Flan-T5 variants.

Method	Flan-T5-Small	Flan-T5-Base	Flan-T5-Large	Avg.
Few-shot	27.40	41.40	57.60	42.13
Few-shot + CoT	26.40	20.80	26.40	24.53
KA. Few-shot + CoT	30.00	46.00	54.00	43.33
Fine-tuning (LoRA)	28.00	41.20	54.80	41.33
self-KD	20.80	32.80	50.80	34.80
prompt-KD	21.40	38.60	47.20	35.73
BOLT	27.40	41.40	57.60	34.13
RTKD	27.00	37.60	41.40	35.33
ours w/o dy	27.60	39.80	53.60	40.33
ours w/o RAG	27.60	43.40	52.20	41.07
ours	27.80	45.60	60.00	**44.47**

Table 18: StrategyQA accuracy (%) of Flan-T5 variants.

Method	Flan-T5-Small	Flan-T5-Base	Flan-T5-Large	Avg.
Zero-shot	50.95	50.07	51.67	50.89
Few-shot	49.20	38.14	56.91	48.08
Few-shot + CoT	51.24	51.53	51.38	51.38
KA. Few-shot + CoT	50.07	51.38	48.18	49.88
Fine-tuning (LoRA)	50.66	21.40	53.42	41.83
self-KD	51.67	54.59	64.63	56.96
prompt-KD	51.82	55.31	52.69	53.27
BOLT	49.20	38.14	56.91	48.08
RTKD	51.38	46.58	38.14	45.37
ours w/o dy	46.87	52.69	51.09	50.22
ours w/o RAG	51.67	50.22	49.20	50.36
ours	48.47	51.97	53.28	**51.24**

Table 19: Effect of Training Data Size on Accuracy (Flan-T5-Large, %).

Setting	Med	Openbook	Strategy	Overall
10% Training Data				
Prompt-KD	23.88	43.80	54.59	36.50
DCRD (ours)	**24.74**	**49.80**	49.49	**36.75**
30% Training Data				
Prompt-KD	22.15	39.80	54.59	34.79
DCRD (ours)	**29.54**	**53.40**	50.36	**40.20**
100% Training Data				
Prompt-KD	22.70	43.60	58.22	36.87
DCRD (ours)	**30.24**	**52.20**	51.82	**40.73**

Overall, these results corroborate the main experimental findings and demonstrate that the improvements of our framework arise from robust gains across multiple datasets rather than from isolated benchmark effects.

B.2 Intermediate Ablation Results of DCRD

This subsection reports supplementary ablation results of DCRD that are not included in the main tables due to space constraints. These results provide fine-grained performance breakdowns across model scales and datasets, supporting the component-level analyses presented in the main paper. All results are evaluated under the same experimental settings as the full model.

B.3 Performance under Different Cross-Domain Imbalance Levels (Flan-T5-Large)

This subsection reports supplementary per-dataset results under different cross-domain imbalance levels using the Flan-T5-Large backbone. While the main paper focuses on overall accuracy trends, these results provide a finer-grained view of how different imbalance mitigation strategies behave across individual datasets. All imbalance settings follow the same construction protocol described in Section 4.4.

B.4 Impact of Training Data Size (Supplementary Results)

This subsection reports supplementary results on the effect of training data size on model accuracy. We evaluate DCRD and a representative distillation baseline (Prompt-KD) under different data availability settings, while keeping the model backbone and optimization settings fixed. These results provide additional evidence on data efficiency and training robustness, complementing the main experimental findings.

B.5 Qualitative Expert Evaluation

Due to space limitations, we provide a complementary qualitative analysis to assess the reasoning quality of different multi-teacher distillation methods. Three domain experts independently evaluate the outputs of each method on 10 randomly sampled QA instances from unseen domains. The experts are instructed to judge (i) answer correctness and (ii) overall reasoning quality, without access to model identities.

For each QA instance, expert preferences are aggregated by majority voting, and results are reported as win rates. As shown in Table **??**, DCRD is consistently preferred over teacher ensemble and MoE-style routing in terms of both correctness and reasoning coherence. These qualitative results align with the quantitative improvements reported in the main experiments and further support the robustness of DCRD.

C Supplementary Experimental Details

This appendix provides detailed implementation settings for all experiments reported in the paper, enabling reproducibility.

C.1 Datasets and Teachers

We consider three heterogeneous knowledge-intensive QA benchmarks: MedQA-USMLE, OpenbookQA, and StrategyQA. Each dataset is associated with a domain-specialized teacher model that provides answer explanations. All teacher-generated data are stored in JSON format and loaded course-wise during training.

C.2 Student Model and Optimization

The student model is initialized from `Flan-T5-Large`. Training is conducted using the AdamW optimizer with a learning rate of 3×10^{-5}. The base batch size is set to 4, and gradient accumulation is dynamically adjusted between 4 and 12 steps depending on the learning state of each course. Multi-GPU training is enabled via data parallelism when available.

C.3 Retrieval-Augmented Distillation

Retrieval-Augmented Generation (RAG) is integrated into the distillation process. For MedQA-USMLE, external knowledge is retrieved from PubMed using the Entrez API, while Wikipedia is used for OpenbookQA and StrategyQA. For each input question, the top-K retrieved documents are concatenated with the question and teacher explanation to form the student input, where $K = 3$ by default.

C.4 Dynamic Curriculum Scheduling

Training proceeds in global iterations. At each iteration, a course is selected according to a weakness-driven sampling distribution based on recent pseudo-accuracy. Courses with lower recent performance are sampled more frequently. For each selected course, both the number of local epochs and the gradient accumulation steps are dynamically adjusted according to the student's recent learning state.

C.5 Pseudo-Accuracy Tracking and Early Stopping

Pseudo-accuracy is computed as $1 - \mathcal{L}_{\text{KD}}$ and tracked separately for each course using a moving window. Training is terminated early if either: (i) all courses exceed a predefined pseudo-accuracy threshold (0.90), or (ii) the loss curves of all courses enter a plateau phase, detected by monitoring recent loss variation. This dual stopping criterion improves training stability and prevents overfitting.

```
 You are a careful reasoning assistant. The following are multiple-choice
questions about medical knowledge.
Use the following retrieved knowledge to help answer the question.
Knowledge: {knowledge_prompt}
Question: {q}
Options:
A. {options[0]}
B. {options[1]}
C. {options[2]}
D. {options[3]}
Please answer with only the letter of the correct option (A, B, C, or D).
Answer:
```

Figure 7: Testing prompt for MedQA-USMLE with retrieval-augmented inference.

```
 You are a careful reasoning assistant. The following are multiple-choice
questions about commonsense knowledge. Output a single option as the final
answer.
Knowledge: {knowledge_prompt}
Question: {q}
Options:
A. {options[0]}
B. {options[1]}
C. {options[2]}
D. {options[3]}
Please answer with only the letter of the correct option (A, B, C, or D).
Answer:
```

Figure 8: Testing prompt for OpenBookQA with retrieval-augmented inference.

C.6 Logging and Visualization

Course-wise loss values and pseudo-accuracy trajectories are logged throughout training. We additionally provide visualizations of loss curves, learning intensity dynamics, and weak-course rankings, which are used for analysis in the supplementary figures.

D Prompts

D.1 Evaluation Prompts

For reproducibility, we report the exact prompts used during the evaluation phase for different datasets. All prompts follow a retrieval-augmented multiple-choice format and require the model to output *only the option letter*, ensuring consistent accuracy measurement.

MedQA-USMLE. For medical-domain questions, we adopt a knowledge-grounded prompt emphasizing clinical reasoning: Table 7.

OpenBookQA. For elementary science and commonsense reasoning, we use a similar format without domain-specific constraints:Table 8.

StrategyQA. For binary decision questions requiring multi-step reasoning, we adopt a two-option prompt:Table 9. All prompts are applied consistently across baselines and our method, with the same retrieved knowledge and decoding constraints.

D.2 Training Prompt

During training, the student model is supervised using teacher-generated explanations and answers, together with retrieved external knowledge. The input prompt is formulated as follows:

```
 You are a careful reasoning assistant. The following are multiple-choice
questions. Output a single option as the final answer.
Knowledge: {knowledge_prompt}
Question: {q}
Options:
A: {options[0]}
B: {options[1]}
Please answer with only the letter of the correct option (A or B).
Answer:
```

Figure 9: Testing prompt for StrategyQA with retrieval-augmented inference.

Instruction: You are a careful reasoning assistant. Use the following retrieved knowledge to answer the question.

Knowledge:
{*Retrieved Documents*}

Question: {*Question*}

Teacher Explanation: {*Teacher-generated Explanation*}

Answer: Answer with the letter only (A/B/C/D).

Figure 10: Training prompt used for retrieval-augmented distillation across all datasets.

This prompt encourages the student model to align its reasoning with both retrieved external knowledge and teacher-provided rationales, enabling effective retrieval-augmented knowledge distillation.

D.3 Few-shot Prompts for Baseline Methods

For fair comparison, we evaluate several baseline methods under a few-shot prompting setting. Following prior work, each prompt consists of multiple question–answer exemplars concatenated before the target query. We report the exact prompts used for different datasets below.

E Algorithm Details

The detailed procedure of DCRD is described in Algorithm 14.

Question	Explanation	Answer
A 3-month-old infant died suddenly during sleep, and autopsy revealed no clear cause of death. Which precaution most likely could have prevented the death?	Detailed Explanations for Medical Questions Question 1: Rib Dysfunction in a Marathon Runner **Question:** A 22-year-old male marathon runner presents with right-sided rib pain during long-distance running. Physical exam shows normal heart and lung findings, but an exhalation dysfunction at ribs 4-5 on the right. Which muscle would be most useful in correcting this dysfunction using a direct method? **Answer:** (C) pectoralis minor **Step-by-step explanation:** 1. **Understanding the dysfunction:** The patient has an exhalation dysfunction at ribs 4-5 on the right. This indicates restricted movement during exhalation, likely causing pain during the repetitive rib movement associated with long-distance running. 2. **Anatomical considerations:** To address rib dysfunction directly, we need to identify muscles that attach to the affected ribs (ribs 4-5 in this case). 3. **Analyzing the options:** - (A) Anterior scalene: Primarily attaches to the cervical vertebrae and first rib, not ribs 4-5. - (B) Latissimus dorsi: Primarily attaches to the vertebrae, iliac crest, and lower ribs (7-12), not ribs 4-5. - (C) Pectoralis minor: Originates from the outer surfaces of the 3rd to 5th ribs, directly attaching to the affected area. - (D) Quadratus lumborum: Attaches to the iliac crest and lumbar vertebrae, with no direct attachment to the ribs. 4. **Respiratory function:** The pectoralis minor acts as an accessory respiratory muscle that can influence rib movement during both inhalation and exhalation. When addressing rib dysfunction, targeting muscles that directly attach to the affected ribs is most effective. 5. **Conclusion:** Among the options, only the pectoralis minor has direct attachment to ribs 4-5, making it the most appropriate muscle to target for correcting this specific rib dysfunction. Question 2: Sacral Dysfunction Diagnosis **Question:** A 36-year-old male with low back pain related to climbing in and out of his truck for work. Prone examination reveals a deep sacral sulcus on the left, a posterior inferior lateral angle (ILA) on the right, and a freely springing lumbosacral junction. What is the most likely diagnosis? **Answer:** (D) right-on-right sacral torsion **Step-by-step explanation:** 1. **Understanding sacral torsion mechanics:** - Sacral torsion involves rotation around an oblique axis - The side with the deep sacral sulcus indicates the direction of rotation - The side with the posterior ILA indicates the direction of sidebend - A freely springing lumbosacral junction indicates no significant flexion or extension component 2. **Analyzing the examination findings:** - Deep sacral sulcus on the left → rotation toward the left - Posterior ILA on the right → sidebend away from the right (toward the left) - Freely springing lumbosacral junction → no flexion or extension component 3. **Interpreting the torsion pattern:** - When the sacrum rotates left and sidebends right, this is described as r̈ight-on-right ẗorsion - R̈ight-on-right ẗmeans the right side of the sacrum moves forward (rotation to the left) while the sacrum sidebends to the right 4. **Evaluating other options:** - (A) Left-on-left sacral torsion would have deep sulcus on the right and posterior ILA on the left - (B) Left-on-right sacral torsion would have deep sulcus on the right and posterior ILA on the left - (C) Right unilateral sacral flexion would have a deep sacral sulcus on the right and possibly a restricted spring test 5. **Conclusion:** The findings (deep left sulcus, right posterior ILA, free spring) are consistent with a right-on-right sacral torsion, which is the most likely diagnosis. Question 3: Sudden Infant Death Prevention **Question:** A 3-month-old baby died suddenly during sleep with no determined cause at autopsy. Which precaution could have prevented this death? **Answer:** (A) Placing the infant in a supine position on a firm mattress while sleeping **Step-by-step explanation:** 1. **Understanding the scenario:** The infant's sudden death during sleep with no apparent cause is characteristic of Sudden Infant Death Syndrome (SIDS), which typically affects infants between 1-4 months of age. 2. **Identifying SIDS risk factors:** - Sleeping in the prone (stomach) position - Sleeping on soft surfaces - Overheating - Loose bedding covering the infant's face - Bed sharing with adults - Exposure to tobacco smoke 3. **Analyzing the preventive options:** - (A) Supine position on firm mattress: This is the single most effective intervention to reduce SIDS risk - (B) Keeping infant covered with high room temperature: This increases overheating risk, which is a SIDS risk factor - (C) Position-maintaining device: Not recommended and may increase risk - (D) Avoiding pacifier use: Actually, pacifier use during sleep is associated with reduced SIDS risk 4. **Evidence for supine positioning:** - Since the American Academy of Pediatrics began recommending supine sleeping in 1992, SIDS rates have decreased by approximately 50% - The mechanism is believed to be related to reduced risk of airway obstruction and overheating - Supine position allows for better air exchange and easier arousal 5. **Additional preventive measures:** - Using a firm sleep surface - Avoiding loose bedding, pillows, and toys in the sleep area - Room-sharing without bed-sharing - Avoiding smoke exposure during pregnancy and after birth - Avoiding overheating 6. **Conclusion:** Placing the infant in a supine position on a firm mattress is the most effective precaution that could have prevented this death, as it addresses the primary modifiable risk factor for SIDS."	A
A 23-year-old pregnant woman at 22 weeks gestation presents with burning upon urination. She states it started 1 day ago and has been worsening despite drinking more water and taking cranberry extract. She otherwise feels well and is followed by a doctor for her pregnancy. Her temperature is 97.7°F (36.5°C), blood pressure is 122/77 mmHg, pulse is 80/min, respirations are 19/min, and oxygen saturation is 98% on room air. Physical exam is notable for an absence of costovertebral angle tenderness and a gravid uterus. Which of the following is the best treatment for this patient?",	Detailed Explanation for Urinary Tract Infection in Pregnancy Question AnalysisThe case involves a 22-week pregnant woman with symptoms of a urinary tract infection (UTI) - burning upon urination that has worsened over 24 hours despite conservative measures. She has no systemic symptoms (normal temperature, no costovertebral angle tenderness), suggesting an uncomplicated lower UTI rather than pyelonephritis. Answer EvaluationLet's analyze each treatment option: (A) Ampicillin- Ampicillin was previously used for UTIs but has fallen out of favor due to high resistance rates (often >30% in E. coli)- Not considered first-line for uncomplicated UTIs in current guidelines- Would require culture and sensitivities to confirm susceptibility (B) Ceftriaxone- Third-generation cephalosporin typically reserved for more serious infections- Used for pyelonephritis or when resistance is a concern- Not first-line for uncomplicated lower UTI- Requires intravenous administration, which is unnecessary for this presentation(C) Doxycycline- Tetracycline class antibiotic- Contraindicated in pregnancy after 14-15 weeks (second and third trimesters) - Can cause fetal bone abnormalities and permanent tooth discoloration- Absolutely contraindicated in this 22-week pregnant patient (D) Nitrofurantoin- First-line treatment for uncomplicated UTIs in non-pregnant adults- Considered appropriate for treatment of uncomplicated UTIs in pregnancy- Advantages: - Excellent urinary concentration - Low resistance rates - Minimal systemic absorption - Generally considered safe in pregnancy (except at term)- Considerations: - Avoid near term (36+ weeks) due to risk of hemolytic anemia in neonates with G6PD deficiency - Some guidelines suggest avoiding in first trimester, though evidence is limited - Appropriate for this patient at 22 weeks gestation Conclusion Nitrofurantoin is the most appropriate treatment choice for this patient. It's effective for uncomplicated UTIs, has good safety profile in pregnancy at this gestational age, and doesn't have the contraindications that the other options possess. The absence of systemic symptoms and normal vital signs suggest this is an uncomplicated lower UTI, not requiring broader-spectrum antibiotics like ceftriaxone, and definitely not contraindicated tetracyclines like doxycycline.Answer: (D) Nitrofurantoin	D

Table 20: Representative examples from the MedQA-USMLE dataset used in our experiments.

Few-shot Prompt for MedQA-USMLE (Multiple-choice, 4 options)

Instruction: You are a careful reasoning assistant. The following are multiple-choice questions about medical knowledge.

Example 1:
A 32-year-old woman with bipolar disorder visits her gynecologist because she believes she is pregnant.
...
Which of the following is the next best step in the management of this patient?
A. Taper lithium and administer valproate
B. Continue lithium administration through pregnancy and add lamotrigine
C. Taper lithium and administer carbamazepine
D. Taper lithium and provide a prescription for clonazepam as needed
Answer: D

Example 2:
A 22-year-old man is brought to the emergency department after falling down stairs.
...
Which neurotransmitter is predominantly blocked by the administered anesthetic?
A. Glutamate
B. Norepinephrine
C. Endorphin
D. Gamma-aminobutyric acid
Answer: A

(Additional medical exemplars omitted for brevity)

Test Question:
Question: {}
(A) {}
(B) {}
(C) {}
(D) {}
Please answer with only the letter of the correct option (A, B, C, or D).

Figure 11: Few-shot prompt used for MedQA-USMLE baseline evaluation.

Table 21: Example questions from OpenBookQA.

Question	Ans.	Explanation
You can make a telescope with a	D	A basic refracting telescope requires a long hollow tube to hold lenses at a fixed distance and block stray light. A mailing tube provides the correct cylindrical structure for assembling a simple telescope, whereas a straw is too narrow, glass alone is insufficient without structure, and a candle has no optical function.
Stars are	C	Stars are massive balls of gas composed mainly of hydrogen and helium. Through nuclear fusion, they release enormous amounts of energy as light and heat. Unlike vague descriptions such as "lights in the sky," this option correctly explains both their composition and their distant astronomical nature.
When food is reduced in the stomach	C	The phrase refers to digestion, where food is physically and chemically broken down inside the stomach. Gastric acids and digestive enzymes deconstruct nutrients into smaller molecules that can later be absorbed by the body. The other options only use the word "digest" metaphorically or irrelevantly.
When standing miles away from Mount Rushmore	D	Due to visual perspective, distant objects appear smaller to the human eye. Photographs often enlarge or emphasize landmarks through camera angles and zoom lenses, making Mount Rushmore appear larger in images than when viewed from far away in person.
The sun is responsible for	D	Plants rely on sunlight for photosynthesis, which powers their entire life cycle including sprouting, blooming, and eventual wilting. The sun is therefore directly responsible for plant growth and development, unlike the unrelated biological or behavioral processes described in the other options.

Few-shot Prompt for OpenBookQA (Commonsense Reasoning)

Instruction: You are a careful reasoning assistant. The following are multiple-choice questions (with answers) about commonsense knowledge.

The sun is responsible for
A. puppies learning new tricks
B. children growing up and getting old
C. flowers wilting in a vase
D. plants sprouting, blooming and wilting
Answer: D

When standing miles away from Mount Rushmore
A. the mountains seem very close
B. the mountains are boring
C. the mountains look the same as from up close
D. the mountains seem smaller than in photographs
Answer: D

(Additional commonsense exemplars omitted for brevity)

Test Question:
Question: {}
(A) {}
(B) {}
(C) {}
(D) {}
Return format: a single uppercase letter only (A, B, C, or D).

Figure 12: Few-shot prompt used for OpenBookQA baseline evaluation.

Few-shot Prompt for StrategyQA (Binary Choice)

Instruction: You are a careful reasoning assistant. The following are multiple-choice questions (with answers).

Is mixed martial arts totally original from Roman Colosseum games?
A: yes
B: no
Answer: B

Is the cuisine of Hawaii suitable for a vegan?
A: yes
B: no
Answer: B

(Additional yes/no exemplars omitted for brevity)

Test Question:
Question: {}
(A) {}
(B) {}
Return format: a single uppercase letter only (A or B).

Figure 13: Few-shot prompt used for StrategyQA baseline evaluation.

Figure 14: Dynamic Curriculum-RAG Distillation (DCRD)

Input: Courses $\mathcal{C} = \{C_1, \ldots, C_N\}$, student model S, teacher models $\{T_i\}$, retriever $\mathcal{R}$
Parameters: threshold τ, base epochs E_0, min/max accumulation MIN_ACC, MAX_ACC
Output: Trained student model S

Initialize moving averages $\overline{\text{Acc}_i^*} \leftarrow 0$ for all C_i
Let training step $t \leftarrow 0$
while training not converged **do**
  $t \leftarrow t + 1$
  Compute course sampling weights
  for each course C_i **do**
    $W_i \leftarrow \max(0.01,\ \tau - \overline{\text{Acc}_i^*})$
  end for
  Sample course $C_i \sim W$
  RAG-enhanced distillation
  Sample question Q from C_i
  Retrieve documents $D \leftarrow \mathcal{R}(Q)$
  Construct input $X \leftarrow \text{Prefix} \parallel D \parallel Q \parallel E_{T_i}$
  Compute distillation loss $\mathcal{L}_{\text{KD}}$
  Update student model S
  Update learning state
  $\text{Acc}_i^* \leftarrow \max(0,\ 1 - \mathcal{L}_{\text{KD}})$
  Update moving average $\overline{\text{Acc}_i^*}$
  Adaptive training intensity
  Set local epochs $E_i \leftarrow \max(1,\ E_0 + \lfloor \tau - \overline{\text{Acc}_i^*} \rfloor)$
  Set gradient accumulation
  $\text{GradAcc}_i \leftarrow \min(\text{MAX_ACC},\ \max(\text{MIN_ACC},\ \lfloor \text{MIN_ACC} + (\tau - \overline{\text{Acc}_i^*}) \rfloor))$
  Train S on C_i for E_i epochs with GradAcc_i
end while
return student model S

Question	Explanation	Answer
Is Mixed martial arts totally original from Roman Colosseum games?	First, we need to understand what mixed martial arts (MMA) is. MMA is a sport that involves various martial arts disciplines, such as boxing, wrestling, jiu-jitsu, and Muay Thai. Participants in MMA fights engage in striking and grappling techniques, attempting to defeat their opponent.Next, we need to look at the history of the Roman Colosseum games. These were gladiatorial games that took place in ancient Rome, where fighters, often slaves or prisoners, would engage in combat for the entertainment of the audience. The fighters would use swords, shields, and other weapons to defeat their opponents.Based on this information, we can conclude that MMA is not totally original from the Roman Colosseum games. While both involve combat and entertainment, the fighting techniques used in MMA are very different from those used in the Colosseum. MMA incorporates a variety of martial arts disciplines, while the Colosseum games were focused on swordsmanship and gladiatorial combat. Additionally, the rules and regulations of MMA are much stricter than those of the Colosseum games, which often resulted in death or serious injury for the fighters.Therefore, the correct answer is (B) no. MMA is not totally original from Roman Colosseum games."	No
Is the cuisine of Hawaii suitable for a vegan?	First, we need to gather some information about the cuisine of Hawaii. Hawaii is known for its fresh seafood, pork, and spam-based dishes, which are not suitable for a vegan diet. Many traditional dishes in Hawaii also include animal products such as eggs, dairy, and honey. Therefore, based on this information, we can conclude that the cuisine of Hawaii is not suitable for a vegan, making answer (B) the correct answer."	No
Is capturing giant squid in natural habitat impossible with no gear?	First of all, giant squid are known to inhabit very deep parts of the ocean, such as the abyssal zone, where there is immense pressure and darkness. Secondly, they are incredibly elusive and difficult to track down in the vastness of the ocean. Thirdly, capturing them would likely require specialized equipment such as submersibles or remotely operated vehicles (ROVs). Without such gear, it would be extremely difficult, if not impossible, to capture giant squid in their natural habitat. Therefore, the answer is (A) yes, capturing giant squid in natural habitat is impossible with no gear.	Yes
Did the Royal Air Force fight in the Boxer Rebellion?	The Boxer Rebellion took place in China in 1900. The Royal Air Force (RAF) was not established until April 1918, well after the Boxer Rebellion had ended. Therefore, it can be concluded that the RAF did not fight in the Boxer Rebellion. The correct answer is (B) no."	No
Do flying fish have good eyesight?	1. Firstly, flying fish are a type of fish that are able to leap out of the water and glide through the air for a short distance.2. Since they are able to fly, it is necessary for them to have good vision in order to navigate and avoid obstacles.3. Many species of fish have very good eyesight, and it is likely that flying fish are no exception.4. Therefore, the answer to the question is most likely yes, flying fish have good eyesight.	Yes
Does a Starbucks passion tea have ginger in it?	First, we need to understand what a Starbucks passion tea is. A passion tea is a type of iced tea that is made from hibiscus flowers, lemongrass, and various fruit flavors. Now, we need to consider whether ginger is typically a component of a passion tea. While ginger is sometimes used in teas for its flavor or health benefits, it is not a typical ingredient in a passion tea. Therefore, the answer to the question is (B) no, a Starbucks passion tea does not have ginger in it."	No
Would Kurt Cobain have benefited from Project Semicolon?	Step 1: What is Project Semicolon? Project Semicolon is a nonprofit organization that aims to raise awareness about mental health and suicide prevention. It encourages individuals to get the support they need by adopting the semicolon as a symbol of hope, love, and continuity in the face of adversity. Step 2: Who is Kurt Cobain? Kurt Cobain was the lead vocalist and guitarist of the American alternative rock band Nirvana. He struggled with depression, chronic pain, and drug addiction throughout his life and died by suicide at the age of 27 in 1994.Step 3: Could Kurt Cobain have benefited from Project Semicolon? Yes, Kurt Cobain could have benefited from Project Semicolon. Project Semicolon provides resources and support for individuals struggling with mental health issues and suicidal thoughts. It promotes the idea that life is worth living and encourages individuals to seek help and find hope. Had Kurt Cobain been aware of Project Semicolon, he may have been able to find the help he needed to cope with his mental health struggles and possibly prevent his suicide. Therefore, the correct answer is (A) Yes, Kurt Cobain could have benefited from Project Semicolon."	yes

Table 22: Representative examples from the StrategyQA dataset used in our experiments.