

000 BEYOND MEAN SCORES: FACTOR MODELS FOR 001 RELIABLE AND EFFICIENT AI EVALUATION 002

003 **Anonymous authors**

004 Paper under double-blind review

005 ABSTRACT

006 Generative AI evaluation relies heavily on benchmark leaderboards, which rank
007 models according to their mean score on a given benchmark. In this paper, we
008 show that these one-number metrics often obscure the multidimensional structure
009 of model capabilities. We propose a factor model approach that decomposes
010 model-item performance into interpretable latent constructs. We apply the
011 modeling approach to examine a novel data set constructed from the Huggingface
012 Open LLM Leaderboard containing item responses from 4,416 language models
013 evaluated across 21,176 questions from six benchmarks. Our analysis reveals two
014 key findings. (i) First, benchmarks contain distinct, sometimes negatively corre-
015 lated constructs that mean scores conflate—models with identical averages can
016 excel at entirely different capabilities. This makes mean scores uninformative—or
017 even misleading—measures of model capabilities. We propose disaggregated
018 alternatives based on the factor structure. (ii) Second, we demonstrate that the
019 factor structure enables efficient estimation of full-benchmark and disaggregated
020 factor-level mean scores. By identifying the most informative questions, we can
021 reduce evaluation costs while preserving model rankings. These results establish
022 factor models as a principled framework for understanding benchmark structure,
023 diagnosing when aggregation obscures meaningful differences, and enabling
024 adaptive evaluation that maximizes information per question.¹

025 1 INTRODUCTION

026 Benchmarks are central to how the AI community measures progress. The standard practice is to
027 summarize model performance with a single “mean score,” which reflects the fraction of items a
028 model was judged as responding to correctly. Mean scores serve as the headline number on bench-
029 mark leaderboards. We argue this practice can be deeply problematic: a single average may mislead
030 by obscuring a multidimensional underlying structure of model capabilities. A key issue is that
031 benchmarks are rarely homogeneous. Intentionally or not, they often combine items that test dif-
032 ferent capabilities, and even a single benchmark item may test a combination of capabilities. Two
033 models with identical mean scores may thus excel on different capability dimensions. For ex-
034 ample, one model might be strong in reasoning but weak in factual recall, while another may have the
035 reverse capability profile. Indeed, in our analysis we find that response correctness across large sub-
036 sets of questions from a given benchmark can be highly *negatively* correlated (see Figure 1). That is,
037 models that perform well on one part of the benchmark are very likely to perform poorly on another.
038 Aggregating these into one number conflates distinct constructs and fails to capture what models ac-
039 tually know and can do. This renders mean scores an inadequate foundation for evaluation.

040 In this paper we propose a more principled approach: Applying factor modeling to directly
041 analyse the multidimensional structure of GenAI model capability that benchmarks implicitly
042 capture. Our approach is grounded in modern psychometrics, which focuses analysis not on mean
043 scores but rather on the *response matrix* of item-level responses across models. In psychometrics,
044 such response matrices have long been analyzed using item response theory (IRT) and factor
045 models to uncover latent constructs. These methods were developed to ensure that standardized
046 exams, such as the GRE (Robin et al., 2014) or personality inventories (Cattell, 1946), measure
047 well-defined underlying traits. Factor models, in particular, reveal clusters of items that load on
048 common capabilities, providing interpretable constructs and enabling measurement beyond simple
049

050 ¹Code is available at <https://anonymous.4open.science/r/factor-model-CE14/>

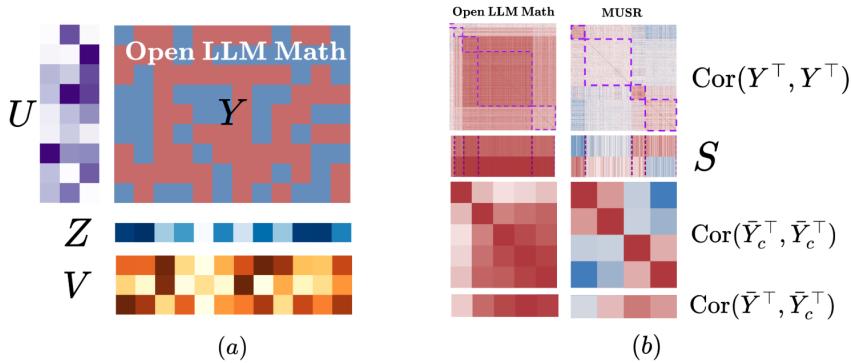


Figure 1: Correlation structure within benchmarks reveals hidden heterogeneity. (a) We show the response matrix for a subset of models and items, decomposed by factor analysis into model factors U , item factors V , and intercepts Z . (b) We show item-item tetrachoric correlations for two benchmarks, Open LLM Math and MUSR. Below each correlation matrix, we present the structure matrix S obtained via factor analysis, along with the cluster-level correlation matrix that summarizes the relationships between cluster mean scores \bar{Y}_c and the overall mean score \bar{Y} . Factor analysis identifies five groups of items in Open LLM Math and four groups of items in MUSR (dashed boxes). For Open LLM Math, the group mean scores are positively correlated with each other, indicating that the benchmark largely measures four related mathematical abilities. Many of them were strongly positive to the overall mean score, indicating that the benchmark-level mean score can be informative about the individual group mean score. In contrast, MUSR exhibits various groups that are negatively correlated, and many of them are weakly (positively or negatively) correlated with the overall mean score, implying the benchmark-level sum score is not too informative about the group mean scores. This illustrates how a single average can obscure and even conflate distinct capabilities: two models with the same mean score may in fact succeed on very different subsets of tasks. Such cases highlight the fragility of mean scores as evaluation metrics and motivate moving toward construct-level analysis via factor models.

aggregates. Yet while these tools have been standard in education and psychology for decades, they have rarely been applied to AI evaluation at scale.

In this paper, we adapt a factor model framework to the setting of large-scale AI benchmarks. Specifically, our analysis focuses on a dataset we curated from the Open LLM Leaderboard, which contains item-level responses from each of 4,416 models to 21,176 questions across 6 benchmarks. We identify a low-rank structure underlying the response matrix, which enables us both to diagnose cases where benchmark mean scores conflate distinct constructs, and also to develop more sample-efficient adaptive evaluation approaches. Our primary contributions are as follows.

1. Revealing benchmark heterogeneity via factor model: The factor model generalizes to held-out models and benchmarks, consistently outperforming baselines and achieving AUCs above 90%, up to 97%. Well-fitted factor models enable the discovery of clusters of items within benchmarks that correspond to distinct, interpretable capabilities. This analysis makes clear why aggregation may mislead: performance on some clusters is strongly negatively correlated with performance on others and with the overall benchmark mean score. Thus, two models with identical mean scores may in fact excel at very different aspects.
2. Developing efficient adaptive evaluation based on factor model: Factor model enables adaptive testing that estimates construct-level scores with far fewer items – reducing the number of items required by the baseline by at least 33.9% and up to 94.3%—cutting evaluation cost while preserving reliability.
3. We present a first systematic attempt to interpret factors discovered by the model. Using the learned loadings together with a modern language model, we generate preliminary semantic labels for the latent constructs. This pipeline goes beyond raw loadings by assigning interpretable meaning and probing item clusters with qualitative examples, offering human-readable explanations of the distinct capabilities they may capture.

In short, we argue that evaluation should move beyond benchmark-level mean scores. Factor models reveal the latent structure of benchmarks, provide more meaningful construct-level measurements, and facilitate efficient evaluation through prediction and adaptive testing. This reframes AI evaluation in line with decades of psychometric practice, offering both better reliability and greater efficiency at leaderboard scale.

2 RELATED WORK

Recent work shows that aggregate scores like mean scores can distort what benchmarks actually reveal about model behavior. Research has shown how leaderboards may fail to assess reliability, hiding persistent failure modes under label noise and ambiguity (Vendrow et al., 2025); and how even small perturbations to preference data can reorder leaderboard rankings (Huang et al., 2025). As Saxon et al. (2024) argues, benchmarks currently function like yardsticks, but they should function more like diagnostic microscopes. Psychometrics offers a principled alternative to such aggregation. The Rasch model (Rasch, 1960) and later two- and three-parameter logistic models (Birnbaum, 1968) provide measurement frameworks that separate item properties from respondent ability, while marginal maximum-likelihood with EM made estimation practical and extensible to multi-dimensional settings (Bock & Aitkin, 1981). Beyond latent-variable approaches, network psychometrics develops graphical formulations that relate closely to (and sometimes subsume) IRT/MIRT (Epskamp et al., 2018). In AI evaluation, ability-oriented measurement has long been advocated as a replacement for task-centric scoring (Hernández-Orallo, 2016).

With the advent of generative AI, these arguments are beginning to gain greater traction. Zhou et al. (2025) propose rubric-engineered, interpretable dimensions with instance-level predictors, but their explanatory power is limited to the pre-defined dimensions and, unlike our work, do not attempt to statistically model the full model-item response matrix. Ruan et al. (2024) introduces observational scaling laws, showing that performance across families lies in a smooth, low-dimensional capability space that enables forecasting. Their study focuses only on prediction and presumes continuous trajectories and, unlike our work, does not disentangle multiple, potentially conflicting constructs. Truong et al. (2025) amortize Rasch-style item-response modeling to reduce evaluation cost, but under a single-trait assumption that does not model multiple constructs. Our work also differs from recent research applying factor modeling and related methods to data already aggregated up to mean scores, such as Burnell et al. (2023), who study latent factors identified from a 29×27 model-benchmark score matrix, and Ren et al. (2024), who conduct PCA on similarly structured data to assess whether safety benchmarks measure constructs distinct from model capabilities. In contrast, our approach (i) models multi-dimensional constructs that can be correlated or even negatively correlated, (ii) identifies them at leaderboard scale by directly fitting the $4,416 \times 21,176$ item response matrix, and (iii) enables new analytic methods—diagnosing when mean scores conflate distinct skills, interpreting factors into human-readable constructs, and powering adaptive testing. These properties complement rubric-based annotation, observational forecasting, and one-dimensional IRT, and are crucial for evaluation at the population scale.

3 METHODS

Factor model for the response matrix. Let $Y_{ij} \sim \text{Bern}(P_{ij})$ denote whether model (“test taker”) $i \in \{1, \dots, N\}$ gives a correct response to question (“item”) $j \in \{1, \dots, M\}$. A factor model (Lord & Novick, 1968; Reckase, 1985) posits that the matrix $P \in [0, 1]^{N \times K}$ has rank K with a logistic link σ with latent parameters (U, V, Z) :

$$P = \sigma(H), \quad H = UV^\top + Z, \quad (1)$$

where $U \in \mathbb{R}^{N \times K}$ are the test taker factors whose covariance is $\Sigma \in \mathbb{R}^{K \times K}$, $V \in \mathbb{R}^{M \times K}$ are the item loadings, and $Z \in \mathbb{R}^M$ are item difficulties (Figure 1). Given the response matrix Y , the parameters can be estimated by a standard inference procedure, such as maximum marginal likelihood, joint maximum likelihood, or Bayesian inference. Because the decomposition is only defined up to a linear transformation, parameter identification is done through centering, whitening, reflection, and rotation constraints (Thurstone, 1935; Anderson & Rubin, 1956; Jöreskog, 1969; Lawley & Maxwell, 1971). Specifically, the estimated parameters $(\hat{U}_{\text{raw}}, \hat{V}_{\text{raw}}, \hat{Z}_{\text{raw}})$ are centered and whitened so that the factor has a zero mean and an identity covariance matrix, while the likelihood remains unchanged:

$$\hat{U}_c = \hat{U}_{\text{raw}} - \bar{U} \implies \hat{Z}_c = \hat{Z}_{\text{raw}} + \hat{V}_{\text{raw}} \bar{U}^\top \quad \text{and} \quad \hat{U}_w = \hat{U}_c L^{-\top} \implies \hat{V}_w = \hat{V} L, \hat{Z}_w = \hat{Z}_c, \quad (2)$$

162 where $\bar{U} = 1/N \sum_i \hat{U}_{\text{raw},i}$ and $\hat{\Sigma} = 1/(N-1) \hat{U}_c^\top \hat{U}_c = LL^\top$. For the reflection constraint,
 163 the loading sign is chosen factor-wise so that the majority of items have positive loading. For the rotation constraint,
 164 we transform the resulting loading to achieve a simple structure, such as one where each loading concentrates on the fewest dimensions, which can be obtained by optimizing the
 165 promax objective (Hendrickson & White, 1964). The factor is adjusted accordingly to maintain the
 166 likelihood. The identified parameters are $(\hat{U}, \hat{V}, \hat{Z})$. Next, we discuss a method to interpret these
 167 estimated parameters.

169 **Understanding item-factor relationships through the structure matrix.** The structure matrix
 170 is a central object in our attempt to interpret the latent structure. To derive the structure matrix,
 171 we begin by considering the thresholded latent response view of the data-generating process. Let
 172 $Y_{ij}^* = H_{ij} + \epsilon_{ij}$, $\epsilon_{ij} \stackrel{\text{iid}}{\sim} \text{Logistic}(0, 1)$. If $Y_{ij} = \mathbf{1}(Y_{ij}^* > 0)$,

$$174 \quad p(Y_{ij} = 1 \mid U_i, Z_j, V_j) = p(\epsilon_{ij} > -Z_j - V_j^\top U_i) = F_{\text{logistic}}(Z_j + V_j^\top U_i) = \sigma(Z_j + V_j^\top U_i). \quad (3)$$

175 The second equality uses the fact that ϵ_{ij} has a cumulative density function F_{logistic} . The last equality
 176 uses $F_{\text{logistic}} = \sigma$. Therefore, the thresholded latent response construction induces exactly the same
 177 Bernoulli model as the initial specification. The latent view is a device that leaves the likelihood
 178 for Y unchanged, yet allows us to treat the right-hand side as a linear factor model with an additive
 179 error whose variance is fixed at $\pi^2/3$. Since ϵ_{ij} is independent of U_i ,

$$181 \quad \text{Cov}(Y_{ij}^*, U_{ik}) = \text{Cov}(H_{ij}, U_{ik}) = \sum_{\ell=1}^K V_{j\ell} \text{Cov}(U_{i\ell}, U_{ik}) = (V_j^\top \Sigma)_k \quad (4)$$

$$184 \quad \text{Var}(Y_{ij}^*) = \text{Var}(H_{ij}) + \text{Var}(\epsilon_{ij}) = V_j^\top \Sigma V_j + \pi^2/3.$$

186 The structure matrix S is defined as the item-factor correlation for the latent response:

$$188 \quad S_{jk} \stackrel{\text{def}}{=} \text{Cor}(Y_{ij}^*, U_{ik}) = \frac{\text{Cov}(Y_{ij}^*, U_{ik})}{\sqrt{\text{Var}(U_{ik}) \text{Var}(Y_{ij}^*)}} = \frac{(V_j^\top \Sigma)_k}{\sqrt{\Sigma_{kk}} \sqrt{V_j^\top \Sigma V_j + \pi^2/3}} \quad (5)$$

191 The row j of the structure matrix indicates how strongly item j correlates with each latent factor. A
 192 set of items that load strongly on the same factor and do not load on other factors can be seen as repre-
 193 senting the behavior domain of that latent dimension. We call these items “unidimensional item.”
 194 Examining the common content of such items helps in interpreting what the latent factor measures
 195 (Gorsuch, 1983; Thurstone, 1947). In practice, there are unfortunately very few unidimensional
 196 items. Besides unidimensional items, it is also helpful to look for a group of items that have similar
 197 loading and form a statistically coherent cluster. Such a cluster can be identified through the struc-
 198 ture matrix by utilizing a clustering algorithm, such as Gaussian mixture models (GMM), where the
 199 number of clusters is determined by minimizing the Bayesian information criterion (BIC). Items are
 200 assigned to a cluster based on the maximum posterior probability under the fitted mixture.

201 **Efficient adaptive evaluation.** Given the estimated item parameters (\hat{V}, \hat{Z}) , the ability-specific fac-
 202 tor can be estimated adaptively to increase the sample efficiency. A popular heuristic for adaptive
 203 item selection is based on the Fisher information. At step t , the Fisher information of item j about the
 204 factor of test taker i is a rank-1 $K \times K$ matrix $\mathbb{I}_{ij}^t = P_{ij}^t (1 - P_{ij}^t) \hat{V}_j \hat{V}_j^\top$, where $P_{ij}^t = \sigma(\hat{V}_j^\top \hat{U}_i^t + \hat{Z}_j)$.
 205 At this step, the informativeness of the item response to the factor can be computed by sequential
 206 optimal design criteria, such as D-optimality, which is the determinant of the accumulated information
 207 matrix (Kiefer, 1959; van der Linden & Glas, 2000):

$$208 \quad j^* = \arg \max_j \det \mathbb{I}_j^{1:t} = \arg \max_j \det \sum_{\tau=1}^t \mathbb{I}_{ij}^\tau. \quad (6)$$

210 The item j^* is then administered to the test taker i , whose response is used to update their corre-
 211 sponding factor via a standard inference procedure such as maximum a posteriori. Per-dimension
 212 measurement reliability is computed as $R_k = 1 - (\hat{\Sigma}_i)_{kk}^2 / \text{Var}(\hat{U}_k)$. The posterior covariance can
 213 be approximated by the inverse Fisher information $\hat{\Sigma}_i \approx (\mathbb{I}_i^{1:t})^{-1}$. We report the mean reliability
 214 $\bar{R} = 1/K \sum_{k=1}^K R_k$. The evaluation procedure terminates when the average reliability reaches a
 215 predetermined level, such as 95%, or when the budget is depleted.

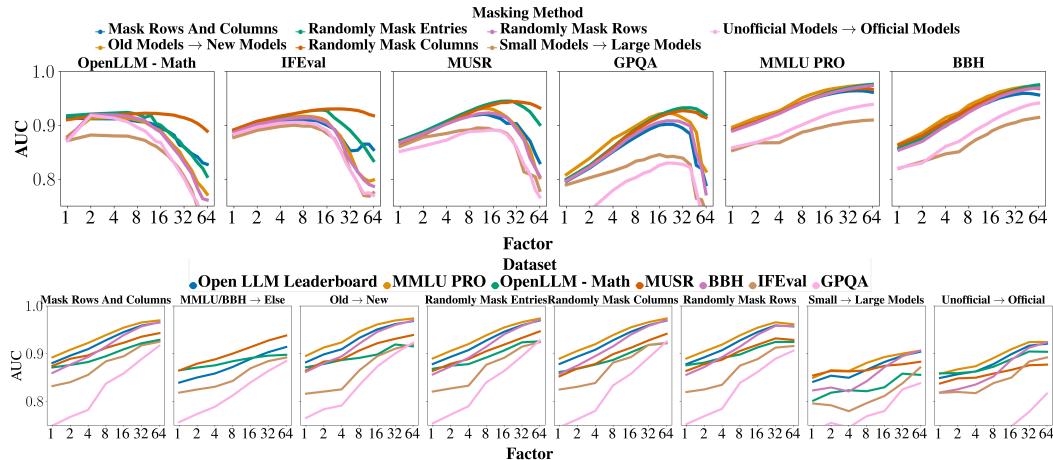


Figure 2: The AUC of the factor model on various ranks and datasets. Different line colors (in top row) or panels (in bottom row) represent different train-test data partition methods for generalization stress tests, such as generalization to unseen language models or questions. **Top row:** The AUC on the test set of the factor model fitted on six individual benchmarks. Each dataset has a different optimal rank. See Table 1 for more details about different data partitioning methods. **Bottom row:** The AUC on the test set of the one-factor model fitted on all six datasets. Higher rank generally generalizes better. (see Subsection C.1 for detailed results)

Tetrachoric correlation. Aside from the factor model, we briefly review the tetrachoric correlation, which is often used in correlation analysis of binary data (Pearson, 1900; Spearman, 1904). Tetrachoric correlation assumes each observed pair $(Y_j, Y_{j'})$ arises from thresholding a latent bivariate normal $(Z_j, Z_{j'})$ with zero mean, unit variances, and correlation ρ . Let (A, B, C, D) denote the 2×2 contingency counts for $(1, 1), (1, 0), (0, 1), (0, 0)$. The implied cell probabilities under thresholds $(\tau_j, \tau_{j'})$ are

$$p_{00} = \Phi(\tau_j, \tau_{j'}; \rho), p_{10} = \Phi(\tau_{j'}) - p_{00}, p_{01} = \Phi(\tau_j) - p_{00}, p_{11} = 1 - \Phi(\tau_j) - \Phi(\tau_{j'}) + p_{00}, \quad (7)$$

where Φ and $\Phi_2(\cdot, \cdot; \rho)$ is the univariate and bivariate normal CDF, respectively. The tetrachoric estimate $\hat{r}_{jj'}^{\text{tet}}$ is the MLE of ρ obtained by maximizing $\ell(\tau_j, \tau_{j'}, \rho) = A \log p_{11} + B \log p_{10} + C \log p_{01} + D \log p_{00}$. Thresholds can be profiled using marginals, e.g. $\hat{\tau}_j = \Phi^{-1}(1 - \hat{\pi}_j)$ with $\hat{\pi}_j = (A + B)/(A + B + C + D)$. Collecting the pairwise MLEs yields the symmetric tetrachoric correlation matrix $R^{\text{tet}} \in \mathbb{R}^{M \times M}$.

4 EXPERIMENTS

We collect response data from the Open LLM Leaderboard (HuggingFace, 2025), a public benchmarking platform maintained by Hugging Face, which evaluates open large language models on a standardized suite of academic and practical tasks to track their capabilities over time. The corpus spans models submitted between 2022 and 2025, covering parameter scales from small models with fewer than 5 billion parameters to large frontier systems with more than 140 billion parameters. In total, our dataset comprises 4,416 distinct language models, each evaluated on 21,176 benchmark questions. The collected response matrix and test takers’ statistics are visualized in Figure 7. The questions come from six widely used evaluation suites: MMLU-Pro (professional-level multi-task language understanding), OpenLLM-Math (mathematical reasoning), MUSR (multi-step reasoning), BBH (Big-Bench Hard), IFEval (instruction-following evaluation), and GPQA (graduate-level problem solving and question answering). The leaderboard data are complete, with no missing entries. We remove questions with a mean accuracy of ≤ 0.01 or ≥ 0.99 across models. For parameter learning in the factor model, we use the standard LBFGS optimizer. We randomly select 80% of the entries of the response matrix for training and test on the rest unless specified otherwise. We use the area under the receiver operating characteristic (AUROC) as our evaluation metric. Unless specified otherwise, we fit one factor model per dataset.

The factor model performance across six benchmarks is shown in Figure 2 and Appendix C.1. Best rank varies: some tasks peak at low ranks (e.g., OpenLLM-Math), while others require larger latent

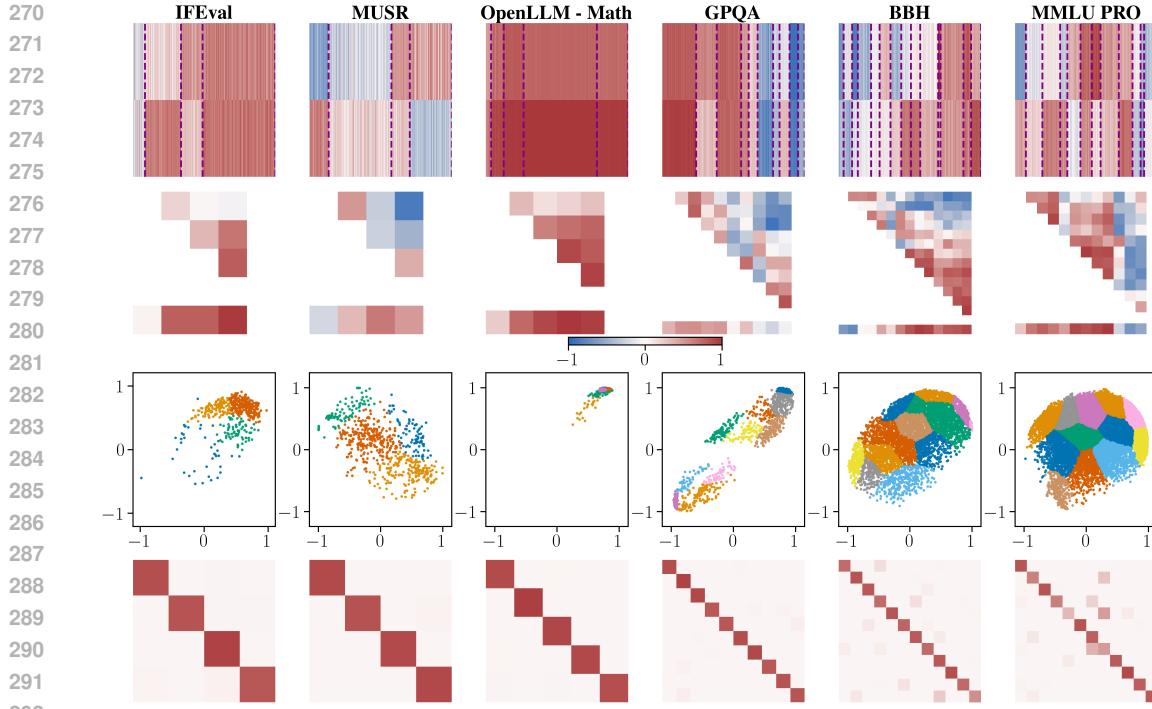


Figure 3: Factor-model constructs within benchmarks. **1st row:** (transposed) structure matrix with items clustered into groups per benchmark by a GMM, where each row is a factor loading, and each column is an item. **2nd row:** correlations between mean scores on clusters, where the last row is the overall benchmark mean score correlation. Across datasets, group mean scores are often only weakly related and can even be negatively correlated with each other and with the overall dataset mean score, indicating that a single average can conflate distinct behaviors. **3rd row:** Visualization of the structure matrix for each dataset. **4th row:** Cluster membership agreement across resampling runs to test cluster robustness. Each row corresponds to a cluster from one subsample of 80% items, with entries showing the percentage of its items that also appear in clusters from another subsample. All overlap matrices are close to identity, showing that cluster structures are stable with perturbation.

spaces. For the best rank across all datasets, the AUC ranges from 92% to 97% for the random mask, indicating an excellent model fit that lays the foundation for further analysis. Additionally, we report a naive baseline for comparison. For each model (row), we compute the mean response across its observed items and use this value as the predicted probability of correctness. This baseline mirrors a common practice in language model evaluation: when only a small fraction of the dataset is available, test performance is approximated by the mean performance on the training set. This baseline yields considerably weaker predictive performance, with AUC values ranging from 53 to 84 for all datasets, consistently lower than those of the factor model.

To assess generalizability under distribution shift, we evaluate the factor model with masking schemes that vary in how train and test entries are selected (see more details in Table 1). Stratified row/column splits shift the marginal distribution over models or items for example, training only on smaller models ($\leq 15B$) and testing on larger ones. We consider four schemes: entry-wise random masking, row holdout (random and shifted), column holdout (random and shifted), and the most challenging row/column block holdout. In row holdout, item parameters are fixed and factors are estimated for new models; in column holdout, model factors are fixed and new items are estimated. Block holdout requires compositional generalization, as both new models and new items must be calibrated from partial responses. In all cases, we use an 80:20 calibration/evaluation split, with 30% of the target block revealed only for factor estimation. In all of these settings, the factor model demonstrates strong generalization capability, implying that the learned factors and loadings can generalize to new models, including the ones that are larger, as well as to new questions and datasets. Last but not least, we fit a single factor model to the entire 6 datasets, rather than one model

324 for each dataset. Figure 2 (bottom row) shows that generally higher rank models perform better on
 325 the test set, which is explained by the fact that they can capture more nuance in the data.
 326

327 4.1 FACTOR MODEL REVEALS BEHAVIOR PATTERNS HIDDEN FROM MEAN SCORES

328 With the well-fitted factor model obtained for each benchmark, we cluster items within each benchmark
 329 by applying a GMM to the structure matrix, selecting the number of clusters by BIC. We
 330 choose to use a rank-2 model for downstream analysis since (1) this model has an excellent fit and
 331 (2) the 2D factor and loading can be visualized easily. We examine two aspects of clustering sta-
 332 bility (Appx. C). First, the chosen number of clusters remains consistent under 10 subsamples at
 333 90% of items. Benchmarks with fewer constructs show near-perfect stability, while larger datasets
 334 exhibit moderate variability but concentrate around a mode (Table 2). Second, we test membership
 335 agreement by comparing cluster overlaps between independent subsamples (see Appendix C for the
 336 definition of the overlap matrix). The resulting matrices is close to identity, indicating that items
 337 consistently group together across runs (Figure 3, fourth row). Overall, clusters are stable under
 338 data perturbation.

339 Having identified the clusters, we compute each models mean score within each cluster (\bar{Y}_c) and the
 340 overall mean score (\bar{Y}), then report the correlations among
 341 them. Figure 3 illustrates this process: the first row
 342 shows the structure matrix with group assignments, the
 343 second row reports correlations between cluster means
 344 and the overall mean, and the third row visualizes the
 345 item structure to highlight the group organization. The
 346 results show that many benchmarks exhibit a heteroge-
 347 neous cluster of items whose means are only weakly cor-
 348 related, while others display clear tension where the two
 349 group means are negatively correlated. When this hap-
 350 pens, the benchmark-level mean score is neither infor-
 351 mative nor accurate about the subgroup score due to het-
 352 erogeneity. Factor models assign a feature vector to each
 353 item (i.e., the structure vector), allowing them to be clus-
 354 tered via a standard clustering algorithm, which helps the
 355 reader better interpret the evaluation result that is missed
 356 by the mean score.

357 We compare two models on MMLU-Pro and show that,
 358 while their overall scores are close (53.1% vs. 54.3%),
 359 they demonstrate different capability profiles (Figure 4)
 360 (Top). This highlights the risk of relying only on mean
 361 benchmark scores: two models with the same average may actually specialize in very different skill
 362 clusters. We further examine how ranking within clusters can diverge from ranking by overall mean
 363 scores. In the MUSR benchmark, the Gaussian mixture model identified four clusters with sizes 83,
 364 190, 87, and 284 (out of 644 items) (Figure 4) (Bottom). Sorting models by their overall mean score
 365 reveals that the 4th cluster, which alone contains about half the dataset ($\sim 44\%$), aligns with the
 366 overall mean: models that rank highly on the overall benchmark also tend to rank highly within this
 367 cluster, although they do not agree perfectly. However, the remaining clusters, which contain 56% of
 368 the dataset, tell a strikingly different story: None of the top-10 models according to the overall mean
 369 is in the top-10 according to the mean score of this 56% subset. Our finding aligns with prior findings
 370 that benchmark heterogeneity strongly influences perceived model effectiveness (Dehghani et al.,
 371 2021). This demonstrates how the overall average can mask severe weaknesses: a model may appear
 372 strong on the overall score, yet perform poorly on a significant fraction of the benchmark.

373 We inspect some questions to understand what the factor model reveals. Here, we highlight two
 374 MUSR questions. Both questions share the same context: “Sarah takes a trowel from the shed and
 375 places it in the front garden, where Emma is helping her. Mr. Brown stays inside the house and
 376 cannot see where the tools have been moved.” They also have the same three choices: backyard,
 377 front garden, and tool shed. The first question asks “Where would Emma most likely look for the
 trowel?” and the second question asks “Where would Mr. Brown most likely look for the trowel?”
 Emma is present and aware that the tool was moved, so she looks in the garden. Mr. Brown, who

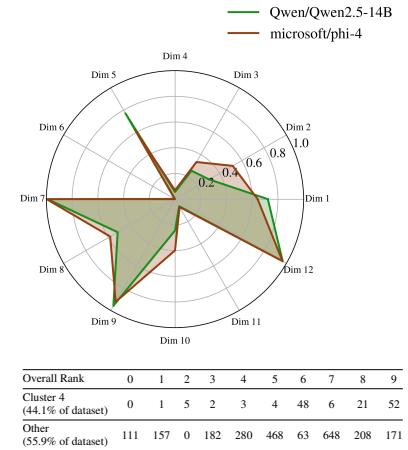


Figure 4: The spider plot compares two models on MMLU-Pro reveals distinct capability profiles despite similar overall scores. The table displays the top 10 model ranks based on the MUSR overall sum score and the two cluster scores.

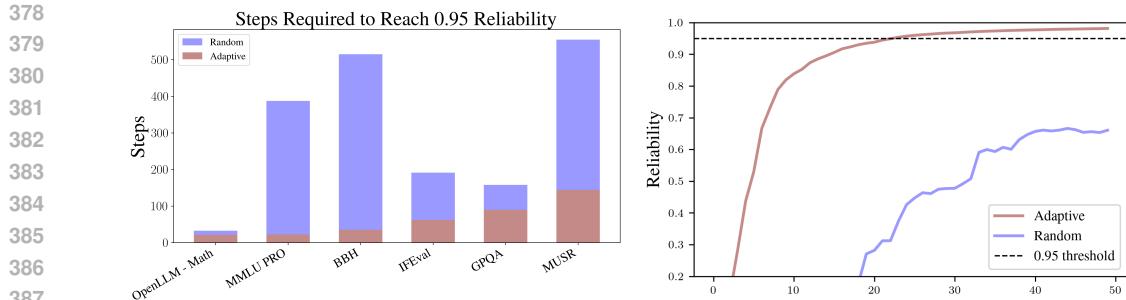


Figure 5: (left) Number of question required to reach average reliability of 0.95 under random vs. adaptive item selection. Adaptive testing consistently requires fewer queries across all benchmarks. (right) Reliability growth for MMLU-Pro. Adaptive selection (red) rapidly achieves high reliability, while random selection (blue) improves more slowly and often falls short of the 0.95 threshold.

is absent, relies on the default assumption that tools remain in the shed. We count model answers across the two questions and find that more than 90% of test takers select the same answer for both questions. They either often retrieve the current location (front garden) or the default storage (tool shed), but rarely adapt depending on the question. We further find that these two questions are highly negatively correlated, with a tetrachoric correlation coefficient response vector of -95.6% . The structure matrix conveys the same message: the structure vectors of the first and second items are $[0.62, -0.34]$ and $[-0.55, 0.42]$, respectively. The sign reversal across both factors places the items in opposite quadrants, indicating opposite loadings, which explain the opposite response pattern for the same test taker. Last but not least, this example also illustrates the advantage of our response-matrix approach over content-based methods: two items may appear textually similar, yet test takers can exhibit sharply different response patterns, which is the center of measurement.

In sum, clustering the factor structure reveals behavioral patterns that are hidden by overall mean scores. While benchmark averages often suggest models perform similarly, factor-based clusters uncover divergent strengths, weaknesses, and even antagonistic response patterns. These results demonstrate that relying solely on mean scores can obscure important aspects of model capability, whereas factor models provide a richer, behavior-driven view of evaluation.

4.2 FACTOR MODEL FACILITATE QUERY EFFICIENT ADAPTIVE TESTING

We have advocated for disaggregating scores from benchmarks and offering a path to interpret them. Nonetheless, given a coherent item pool, we still want to estimate a summary of model performance on this group of items, which can be expensive for a large group. We demonstrate that the factor model can help improve query efficiency in evaluating this item pool via adaptive testing. To illustrate this, we randomly sampled 200 held-out LLMs, for each of which we administered items with learned parameters chosen either randomly or adaptively using Fisher information. Figure 5 (left) shows that adaptive testing consistently requires far fewer items than random selection across benchmarks—reducing the number of items required by the baseline by at least 33.9% and up to 94.3%. The right panel illustrates this pattern on the MMLU-Pro dataset: adaptive selection rapidly drives reliability toward saturation, reaching 95% after 20 steps, while random selection improves much more slowly. Overall, adaptive testing with the factor model achieves high reliability using far fewer queries than random selection. By targeting the most informative items, it reduces evaluation cost by an order of magnitude in some benchmarks while preserving measurement accuracy. This demonstrates that factor-based adaptive testing offers a practical and efficient alternative to random or exhaustive benchmark evaluation.

4.3 INTERPRETATION OF BEHAVIOR DOMAIN FROM FACTOR MODEL

Factors from the factor model represent the underlying construct that governs the behavior of a group of items. This group of items exhibits a coherent response pattern, which helps interpret the observable behavior domain that links to the construct. Below, we attempt to interpret these behavior domains. Our goal here is not to claim a definitive or optimal interpretation, but rather to provide a first systematic attempt at making sense of the discovered factors. Even a preliminary interpretation adds value by moving beyond purely numerical factor loadings and offering a starting

432 point for understanding what capabilities benchmarks implicitly capture. We view this as an initial
 433 step toward a richer interpretive framework that future work can refine and extend.
 434

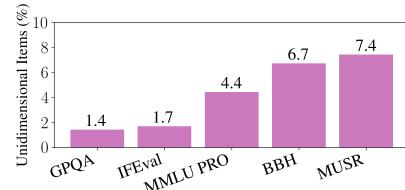
435 For a target dimension, we use the structure matrix to isolate unidimensional items: items with only
 436 one dimension of significant absolute value above a threshold $\tau = 0.2$, and insignificant everywhere
 437 else. Such items are only associated with the targeted factor and form the candidate set, while
 438 negatives are sampled from items with negligible association with the target construct. Ten positive
 439 and thirty negative items are then submitted to GPT-5, which infers the distinguishing content of the
 440 positives to describe the factor’s domain. Three human reviewers then review the inferred domain
 441 from GPT5. For example, applied to MMLU-Pro with $K = 4$, one dimension corresponds to
 442 “concept classification & single-step quantitative application”: the positive items ask you either to
 443 identify the correct conceptual category/definition (e.g., progressive vs. regressive tax) or to apply
 444 one canonical formula with straightforward reasoning (e.g., photon emission rate, spring energy
 445 $W = \frac{1}{2}kx^2$, polymer M_n/M_w , basic heat-transfer correlations). By contrast, the negatives require
 446 multi-step technical calculations, procedural ordering in law, or rote historical facts, going beyond
 447 quick recognition or a single canonical equation.

448 Unidimensional items are rare in benchmarks (often < 10%; see Figure 6), so evaluations based only on them
 449 are noisy even if interpretable. To capture broader constructs, we instead use clusters of statistically coherent
 450 items (e.g., with positive inter-item correlation). For each
 451 cluster, we select 10 items as positives and sample 60 neg-
 452 atives from other clusters. These examples are then given
 453 to GPT-5, which identifies the content that distinguishes
 454 the positives from the negatives and returns a concise la-
 455 bel and explanation, yielding a human-interpretable de-
 456 scription of the cluster. Three human reviewers then re-
 457 view the inferred cluster content from GPT5. In MMLU-
 458 Pro, one group was found to be principle and relationship
 459 reasoning (cause-effect and structural rules). The posi-
 460 tive items ask you to identify the governing principle or
 461 how variables interactas in legal doctrines (hearsay, jurisdiction), ML bagging (bootstrap sampling),
 462 optics trends near the focal point, macro trade-offs (Phillips curve), and degree growth under poly-
 463 nomial compositionrather than doing heavy numeric computation or recalling isolated facts (which
 464 dominate the negatives). For the detailed prompt used (see Subsection D).

465 In summary, the factor model reveals behavior domains by grouping items into coherent groups.
 466 Although unidimensional items can be interpretable, they are too sparse to support robust eval-
 467 uation. Clustering provides a richer alternative, uncovering statistically coherent item groups that can
 468 be mapped to human-interpretable domains. For researchers, this offers a window into the latent
 469 structure of benchmarks. For practitioners, it provides interpretable summaries of model behavior
 470 that go beyond raw scores and can guide more targeted evaluation.

471 5 CONCLUSION, LIMITATIONS, FUTURE WORK

472 Mean scores blur what benchmarks really measure. By treating the response matrix with a low-
 473 to-moderate rank factor model, we recover interpretable constructs, accurately impute held-out re-
 474 sponses for unseen models and items, and run adaptive tests that reach high reliability with far fewer
 475 queries. Disaggregating by constructs reveals when identical averages mask opposing strengths,
 476 providing more accurate reporting and more targeted evaluation. Our study inherits constraints
 477 from leaderboard data: binary grading and dataset-specific biases. Behavior domain interpre-
 478 tation elicited can be subjective. Several directions can extend this framework. Methodologically,
 479 moving beyond dichotomous scoring to incorporate partial-credit and open-ended judgments will
 480 improve fidelity. Testing for measurement invariance and differential item functioning across lan-
 481 guages, demographics, and prompting styles will enhance the robustness of construct-level eval-
 482 uation. Coupling factor models with item content features could support cold-start evaluation on
 483 unseen benchmarks. Finally, open-source tooling for construct-level reporting, adaptive eval-
 484 uation, and judge-drift modeling would facilitate broader adoption in both research and governance
 485 contexts. Together, these advances will help establish AI evaluation as a reliable and efficient mea-
 surement science for AI.



476 Figure 6: Proportion of unidimensional
 477 items across datasets. About 1 - 7%
 478 of items meet the unidimensionality cri-
 479 terion, corresponding to 53 items in
 480 MUSR, 17 in GPQA, 9 in IFEval, 388
 481 in BBH, and 526 in MMLU-Pro.

486 REFERENCES
487

488 Theodore W. Anderson and Herman Rubin. Statistical inference in factor analysis. In Jerzy Neyman
489 (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*,
490 volume 5, pp. 111–150, Berkeley, 1956. University of California Press.

491 Allan Birnbaum. Some latent trait models and their use in inferring an examinee's ability. In Fred-
492 eric M. Lord and Melvin R. Novick (eds.), *Statistical Theories of Mental Test Scores*, pp. 395–
493 479. AddisonWesley, Reading, MA, 1968. URL https://books.google.com/books?id=k_wnDwAAQBAJ.

494

495 R. Darrell Bock and Murray Aitkin. Marginal maximum likelihood estimation of item pa-
496 rameters: Application of an em algorithm. *Psychometrika*, 46(4):443–459, 1981. doi:
497 10.1007/BF02293801. URL <https://link.springer.com/article/10.1007/BF02293801>.

498

499 500 Ryan Burnell, Han Hao, Andrew RA Conway, and Jose Hernandez Orallo. Revealing the structure
501 of language model capabilities. *arXiv preprint arXiv:2306.10062*, 2023.

502

503 504 Raymond B. Cattell. *Description and Measurement of Personality*. World Book Com-
505 pany, YonkersHudson, NY, 1946. URL <https://www.semanticscholar.org/paper/The-Description-and-Measurement-of-Personality-Cattell/b2db0f2395eaadb0f57e0bdbecf68bac5a3bb274>. Reprinted and discussed in later
506 psychometric literature.

507

508 509 Mostafa Dehghani, Yi Tay, Alexey A. Gritsenko, Zhe Zhao, Neil Houlsby, Fernando Diaz, Donald
510 Metzler, and Oriol Vinyals. The benchmark lottery, July 2021. URL <https://arxiv.org/abs/2107.07002>. arXiv preprint.

511

512 513 Sacha Epskamp, Mijke Rhemtulla, and Denny Borsboom. The gaussian graphical model in
514 cross-sectional and time-series data. *Multivariate Behavioral Research*, 53(4):453–480, 2018.
515 doi: 10.1080/00273171.2018.1454823. URL <https://psycnet.apa.org/record/2018-10249-031>.

516

517 Richard L. Gorsuch. *Factor Analysis*. Lawrence Erlbaum Associates, Hillsdale, NJ, 2 edition, 1983.

518

519 Andrew E. Hendrickson and Paul O. White. Promax: A quick method for rotation to oblique simple
520 structure. *British Journal of Statistical Psychology*, 17(1):65–70, 1964.

521

522 José Hernández-Orallo. Evaluation in artificial intelligence: From task-oriented to ability-
523 oriented measurement. *Artificial Intelligence Review*, 48(3):397–447, 2016. ISSN 0269-
524 2821, 1573-7462. doi: 10.1007/s10462-016-9505-7. URL <https://doi.org/10.1007/s10462-016-9505-7>.

525

526 Jenny Y. Huang, Yunyi Shen, Dennis Wei, and Tamara Broderick. Dropping just a handful of
527 preferences can change top large language model rankings. In *2nd Workshop on Models of Human
Feedback for AI Alignment*, 2025. URL <https://arxiv.org/abs/2508.11847>.

528

529 HuggingFace. Open llm leaderboard. <https://huggingface.co/open-llm-leaderboard>, 2025. Accessed: 2025-09-24.

530

531 Karl G. Jöreskog. A general approach to confirmatory maximum likelihood factor analysis. *Psy-
532 chometrika*, 34(2):183–202, 1969. doi: 10.1007/BF02289343.

533

534 Jack Kiefer. Optimum experimental designs. *Journal of the Royal Statistical Society: Series B
(Methodological)*, 21(2):272–319, 1959.

535

536 D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. Butterworths, London,
537 1971.

538

539 Frederic M. Lord and Melvin R. Novick. *Statistical Theories of Mental Test Scores*. Addison-
Wesley, Reading, MA, 1968.

540 Karl Pearson. Mathematical contributions to the theory of evolution. VII. on the correlation of
 541 characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of*
 542 *London. Series A, Containing Papers of a Mathematical or Physical Character*, 195:1–47, 1900.
 543 doi: 10.1098/rsta.1900.0022.
 544

545 Georg Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. University of
 546 Chicago Press, Chicago, 1960. ISBN 978-0226705538. URL <https://archive.org/details/probabilisticmod0000rasc>. Expanded ed., foreword and afterword by B.
 547 D. Wright.
 548

550 Mark D. Reckase. The difficulty of test items that measure more than one ability. *Applied Psycho-
 551 logical Measurement*, 9(4):401–412, 1985.
 552

553 Richard Ren, Steven Basart, Adam Khoja, Alice Gatti, Long Phan, Xuwang Yin, Mantas Mazeika,
 554 Alexander Pan, Gabriel Mukobi, Ryan Kim, et al. Safetywashing: Do ai safety benchmarks
 555 actually measure safety progress? *Advances in Neural Information Processing Systems*, 37:
 556 68559–68594, 2024.
 557

558 Frédéric Robin, Isaac Bejar, Longjuan Liang, and Frank Rijmen. Dimensionality analyses of the
 559 GRE® revised general test verbal and quantitative measures. Technical Report GREB-RR-14-xx,
 560 Educational Testing Service, Princeton, NJ, 2014. URL <https://files.eric.ed.gov/fulltext/EJ1124807.pdf>.
 561

563 Yangjun Ruan, Chris J. Maddison, and Tatsunori Hashimoto. Observational scaling laws and the
 564 predictability of language model performance. In *Proceedings of the 38th International Con-
 565 ference on Neural Information Processing Systems (NeurIPS)*, 2024. URL <https://arxiv.org/abs/2405.10938>.
 566

568 Michael Saxon, Ari Holtzman, Peter West, William Wang, and Naomi Saphra. Benchmarks as
 569 microscopes: A call for model metrology. *arXiv preprint arXiv:2407.16711*, 2024. URL <https://arxiv.org/pdf/2407.16711.pdf>.
 570

572 Charles Spearman. “general intelligence,” objectively determined and measured. *The American
 573 Journal of Psychology*, 15(2):201–293, 1904. doi: 10.2307/1412107.
 574

575 Louis L. Thurstone. *The Vectors of Mind: Multiple-Factor Analysis for the Isolation of Primary
 576 Traits*. University of Chicago Press, Chicago, 1935.
 577

578 Louis L. Thurstone. *Multiple-Factor Analysis*. University of Chicago Press, Chicago, IL, 1947.
 579

580 Sang Truong, Yuheng Tu, Percy Liang, Bo Li, and Sanmi Koyejo. Reliable and efficient amortized
 581 model-based evaluation. In *International Conference in Machine Learning*, 2025. URL <https://arxiv.org/abs/2503.13335>.
 582

584 Wim J. van der Linden and Cees A.W. Glas (eds.). *Computerized Adaptive Testing: Theory and
 585 Practice*. Kluwer Academic Publishers, Dordrecht, 2000.
 586

587 Joshua Vendrow, Edward Vendrow, Sara Beery, and Aleksander Madry. Do large language model
 588 benchmarks test reliability?, 2025. URL <https://arxiv.org/abs/2502.03461>.
 589

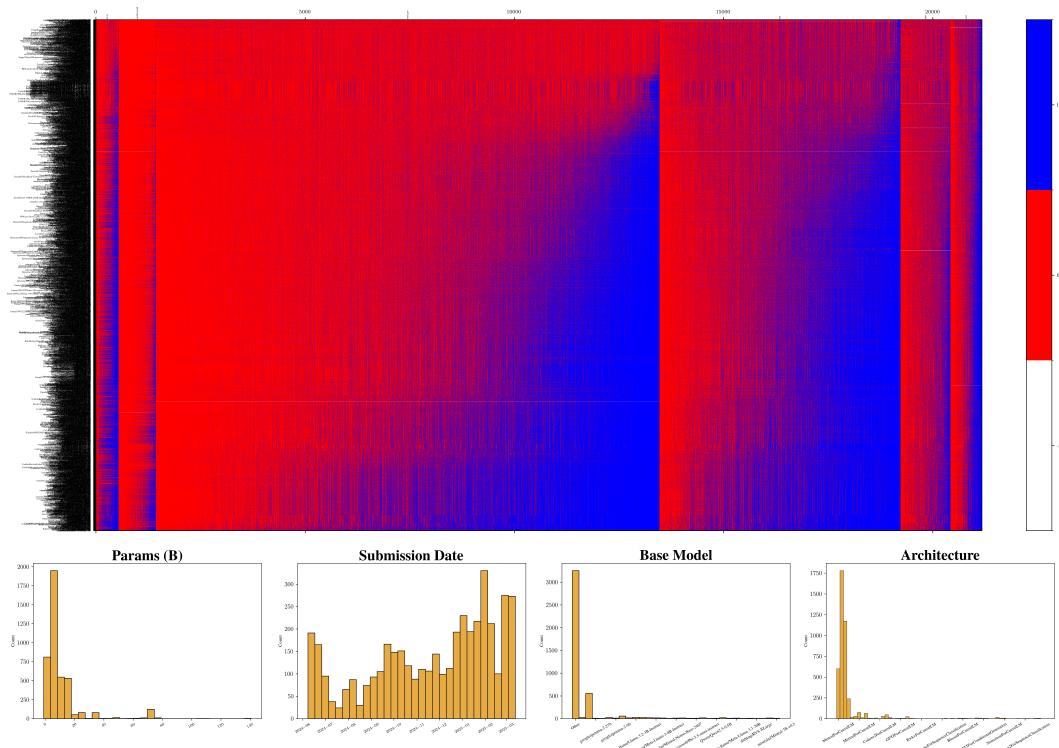
591 Lexin Zhou, Lorenzo Pacchiardi, Omar Shaikh, Chen Zhu, Nathaniel Thiebaut, Alejandro Romero-
 592 Soriano, Aditi Singhal, Yulin Wang, Yilun Du, Sébastien Bubeck, and et al. General scales
 593 unlock ai evaluation with explanatory and predictive power. *arXiv preprint arXiv:2503.06378*,
 2025. URL <https://arxiv.org/abs/2503.06378>.

594 A USE OF LARGE LANGUAGE MODELS (LLMs)

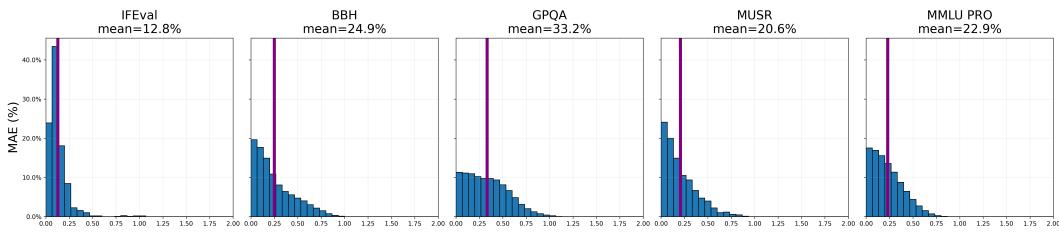
596 We used Large Language Models (LLMs) in the preparation of this work as general-purpose as-
 597 sistant tools for two purposes: (1) polishing the writing, including grammar and readability, and
 598 (2) providing coding assistance. In all cases, outputs generated by LLMs were carefully reviewed,
 599 verified, and modified by the authors before inclusion. The authors take full responsibility for all
 600 content presented in this paper.

601 B ADDITIONAL RESULTS

603 Below we present additional results.



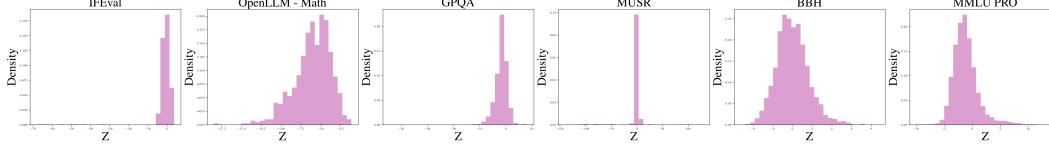
629 Figure 7: (Top) Visualization of the response matrix with 4,416 models and 21,176 questions. The
 630 matrix is sorted first by model mean accuracy (from weaker to stronger models), and within each
 631 benchmark by question mean accuracy (from harder to easier questions). (Bottom) We show the
 632 histogram of test takers' characteristics.



641 Figure 8: Sanity check for the structure matrix S . To validate that the estimated structure matrix
 642 indeed behaves as the correlation between the latent factor and the item response, we compare it with
 643 the approximated empirical Pearson item-factor correlations computed from held-out responses. The
 644 appropriate correlation here is polyserial, but they are expensive to compute, so we approximate with
 645 Pearson. For each item j and factor k , we estimate $\hat{r}_{jk} = \text{Cor}(Y_{\cdot j}, \hat{U}_{\cdot k})$ across models. Each panel
 646 shows the distribution of absolute deviations $|\hat{r}_{jk} - S_{jk}|$ between empirical itemfactor correlations
 647 and model-implied loadings for one benchmark. The concentration of values near zero confirms that
 the structure matrix is close to the empirical correlations, verifying that they behave as expected.

648
649
650
651 Table 1: Masking schemes for evaluating factor model generalization. Each scheme specifies how
652 train/test entries are selected and which parameters are re-estimated.
653

651 Masking Type	652 Train Set	653 Test Set	654 Purpose / Procedure
655 Entry-wise random	656 80% random entries across 657 the full matrix	658 20% random entries	659 Interpolation under missing-at-random; no distribution shift
660 Row holdout (random)	661 80% of models, all items	662 20% of models, all items	663 Generalization to unseen models; freeze V, Z from train set and estimate U for test models using revealed responses
664 Row holdout (shifted)	665 Slice of model population	666 Disjoint slice: small \rightarrow large models older \rightarrow newer releases unofficial \rightarrow official providers	667 Tests transfer of item loadings V under covariate shift; freeze V, Z from train set and estimate U for test models
668 Column holdout (random)	669 All models, 80% of items	670 All models, 20% of items	671 Generalization to unseen items; freeze U from train set and estimate V, Z for test items using revealed responses
672 Column holdout (shifted)	673 Train on some benchmarks	674 Hold out benchmarks: MMLU+BBH (train) GPQA, MUSR (test)	675 Tests transfer of model factors U across domains; freeze U from train set and estimate V, Z for test items
676 Row-column block holdout	677 $R_{\text{tr}} \times C_{\text{tr}}$ (train rows, train cols)	678 $R_{\text{te}} \times C_{\text{te}}$ (test rows, test cols)	679 Most challenging: compositional generalization; estimate U for new models with V, Z fixed, and estimate V, Z for new items with U fixed, then predict the held-out block

675
676 Figure 9: Distribution of Z in different benchmarks
677678

CLUSTER STABILITY AND MEMBERSHIP AGREEMENT

679
680 Let $\mathcal{G} = \{G_a\}_{a=1}^K$ and $\mathcal{H} = \{H_b\}_{b=1}^{K'}$ denote two clusterings (e.g., two subsample runs) over the
681 same benchmark item set (after subsampling). We define the (row-normalized, asymmetric) *overlap*
682 matrix $O \in \mathbb{R}^{K \times K'}$ by

$$683 O_{ab} = \frac{|G_a \cap H_b|}{|G_a|}.$$

$$684$$

685 Each row sums to at most 1 because we subsample 90% of items in each run. A simple *row purity*
686 metric is $\max_b O_{ab} \in [0, 1]$.

687 For each benchmark, we perform 10 runs in which we (i) keep a random 90% subset of items, (ii)
688 fit a GMM to the structure matrix, and (iii) select the number of clusters via BIC. We record the
689 selected K and compute overlap matrices between runs to visualize membership agreement (bold
690 entries in main-text tables mark row-wise maxima). Across the 10 subsamples, the BIC-selected
691 cluster counts are:

692 Benchmark	693 Cluster counts across 10 runs
694 IFEVAL	695 [4, 4, 4, 4, 4, 4, 4, 4, 4, 4]
696 OpenLLM-Math	697 [5, 5, 5, 5, 5, 5, 4, 5, 4]
698 GPQA	699 [11, 10, 10, 9, 9, 11, 9, 9, 9]
MUSR	[2, 2, 5, 4, 4, 5, 5, 4, 2, 5]
BBH	[12, 12, 11, 10, 12, 11, 13, 14, 14, 17]
MMLU-Pro	[12, 12, 11, 10, 10, 11, 12, 11, 14, 12]

700 Table 2: Cluster counts under 10 subsampling runs (90% of items). Smaller datasets (e.g., IFEVAL)
701 are perfectly stable, while larger benchmarks (e.g., BBH, MMLU-PRO) exhibit moderate variability
but still center around a mode.

Membership agreement. For each pair of runs we report O ; row-wise maxima (bold) identify a single best-matching cluster in the comparison run. In the main text, the overlap matrices (e.g., for IFEVAL, OpenLLM-Math, GPQA, MUSR, BBH) show large row purities (many > 0.8), indicating stable membership. Because O is asymmetric (conditioned on G_a), one may also compute the symmetric variant by normalizing columns or by using $\min\{|G_a \cap H_b|/|G_a|, |G_a \cap H_b|/|H_b|\}$; conclusions are qualitatively unchanged.

Takeaways. (1) The chosen number of constructs is stable under subsampling, concentrating around a clear mode for each dataset. (2) Cluster membership is stable: each rows mass concentrates on a single match, supporting our claim that the discovered constructs are reproducible and not artifacts of a particular sample of items.

C.1 TRAIN AND TEST DATA AUC

Table 3: Performance comparison across datasets and masking methods for AUC, Correlation, and Log Probability metrics (Test Split)

Dataset	Masking	K=0	K=1	K=2	K=4	K=8	K=16	K=32	K=64
AUC (x 100)									
Ifeval	rand rand	74.84 \pm 0.00	88.92 \pm 0.00	90.56 \pm 0.00	91.59 \pm 0.00	92.49 \pm 0.00	92.70 \pm 0.00	89.04 \pm 0.00	83.44 \pm 0.00
	randrow randrow	74.44 \pm 0.00	88.92 \pm 0.00	90.59 \pm 0.00	91.03 \pm 0.00	91.39 \pm 0.00	90.45 \pm 0.00	83.39 \pm 0.00	78.65 \pm 0.00
	old new	75.11 \pm 0.00	89.04 \pm 0.00	90.56 \pm 0.00	91.25 \pm 0.00	91.42 \pm 0.00	90.73 \pm 0.00	84.46 \pm 0.00	79.86 \pm 0.00
	small large	74.56 \pm 0.00	87.73 \pm 0.00	89.05 \pm 0.00	89.78 \pm 0.00	89.85 \pm 0.00	88.48 \pm 0.00	81.65 \pm 0.00	77.48 \pm 0.00
	unofficial official	75.01 \pm 0.00	88.33 \pm 0.00	89.65 \pm 0.00	90.40 \pm 0.00	90.54 \pm 0.00	89.05 \pm 0.00	81.66 \pm 0.00	76.96 \pm 0.00
	randcol randcol	74.84 \pm 0.00	89.18 \pm 0.00	90.76 \pm 0.00	91.70 \pm 0.00	92.51 \pm 0.00	93.08 \pm 0.00	92.92 \pm 0.00	91.79 \pm 0.00
Openlm Math	L	76.10 \pm 0.00	88.10 \pm 0.00	89.68 \pm 0.00	90.63 \pm 0.00	90.95 \pm 0.00	90.02 \pm 0.00	85.26 \pm 0.00	85.49 \pm 0.00
	rand rand	80.93 \pm 0.00	91.78 \pm 0.00	92.06 \pm 0.00	92.28 \pm 0.00	92.34 \pm 0.00	90.08 \pm 0.00	85.34 \pm 0.00	80.51 \pm 0.00
	randrow randrow	81.04 \pm 0.00	87.52 \pm 0.00	92.09 \pm 0.00	91.25 \pm 0.00	92.37 \pm 0.00	83.87 \pm 0.00	83.66 \pm 0.00	76.13 \pm 0.00
	old new	79.78 \pm 0.00	87.93 \pm 0.00	91.33 \pm 0.00	91.41 \pm 0.00	90.97 \pm 0.00	88.99 \pm 0.00	83.72 \pm 0.00	77.20 \pm 0.00
	small large	75.65 \pm 0.00	87.16 \pm 0.00	88.19 \pm 0.00	88.03 \pm 0.00	87.56 \pm 0.00	85.44 \pm 0.00	80.54 \pm 0.00	71.10 \pm 0.00
	unofficial official	83.40 \pm 0.00	87.15 \pm 0.00	91.83 \pm 0.00	91.33 \pm 0.00	89.61 \pm 0.00	86.81 \pm 0.00	80.18 \pm 0.00	71.42 \pm 0.00
Mmlu Pro	randcol randcol	80.56 \pm 0.00	91.04 \pm 0.00	91.67 \pm 0.00	91.91 \pm 0.00	92.15 \pm 0.00	92.12 \pm 0.00	91.42 \pm 0.00	88.92 \pm 0.00
	L	79.56 \pm 0.00	91.33 \pm 0.00	91.22 \pm 0.00	91.20 \pm 0.00	90.73 \pm 0.00	89.49 \pm 0.00	86.03 \pm 0.00	82.75 \pm 0.00
	rand rand	66.08 \pm 0.00	89.06 \pm 0.00	90.66 \pm 0.00	92.25 \pm 0.00	94.35 \pm 0.00	95.92 \pm 0.00	96.96 \pm 0.00	97.61 \pm 0.00
	randrow randrow	66.36 \pm 0.00	88.94 \pm 0.00	90.47 \pm 0.00	92.18 \pm 0.00	94.15 \pm 0.00	95.71 \pm 0.00	96.77 \pm 0.00	97.38 \pm 0.00
	old new	67.66 \pm 0.00	89.67 \pm 0.00	91.39 \pm 0.00	92.68 \pm 0.00	95.13 \pm 0.00	96.53 \pm 0.00	97.23 \pm 0.00	97.59 \pm 0.00
	small large	61.52 \pm 0.00	85.31 \pm 0.00	86.74 \pm 0.00	86.77 \pm 0.00	88.74 \pm 0.00	89.89 \pm 0.00	90.60 \pm 0.00	90.98 \pm 0.00
Bbh	unofficial official	68.72 \pm 0.00	85.85 \pm 0.00	86.82 \pm 0.00	88.18 \pm 0.00	90.11 \pm 0.00	91.83 \pm 0.00	93.13 \pm 0.00	93.87 \pm 0.00
	randcol randcol	65.90 \pm 0.00	89.18 \pm 0.00	90.72 \pm 0.00	92.33 \pm 0.00	94.34 \pm 0.00	95.89 \pm 0.00	96.65 \pm 0.00	96.56 \pm 0.00
	L	65.97 \pm 0.00	89.05 \pm 0.00	90.69 \pm 0.00	92.22 \pm 0.00	94.20 \pm 0.00	95.62 \pm 0.00	96.36 \pm 0.00	96.13 \pm 0.00
	rand rand	62.55 \pm 0.00	85.84 \pm 0.00	87.59 \pm 0.00	90.20 \pm 0.00	92.83 \pm 0.00	95.02 \pm 0.00	96.62 \pm 0.00	97.52 \pm 0.00
	randrow randrow	62.41 \pm 0.00	85.39 \pm 0.00	87.00 \pm 0.00	89.90 \pm 0.00	92.42 \pm 0.00	94.64 \pm 0.00	96.22 \pm 0.00	96.94 \pm 0.00
	old new	63.71 \pm 0.00	86.51 \pm 0.00	88.57 \pm 0.00	91.42 \pm 0.00	93.73 \pm 0.00	95.62 \pm 0.00	96.88 \pm 0.00	97.28 \pm 0.00
Gpqa	small large	60.23 \pm 0.00	82.01 \pm 0.00	83.26 \pm 0.00	84.69 \pm 0.00	86.44 \pm 0.00	88.97 \pm 0.00	90.51 \pm 0.00	91.46 \pm 0.00
	unofficial official	63.26 \pm 0.00	82.10 \pm 0.00	83.01 \pm 0.00	86.05 \pm 0.00	88.23 \pm 0.00	90.84 \pm 0.00	92.89 \pm 0.00	94.12 \pm 0.00
	randcol randcol	62.47 \pm 0.00	86.35 \pm 0.00	88.07 \pm 0.00	90.66 \pm 0.00	93.02 \pm 0.00	95.08 \pm 0.00	96.41 \pm 0.00	96.86 \pm 0.00
	L	63.07 \pm 0.00	85.54 \pm 0.00	87.41 \pm 0.00	89.83 \pm 0.00	92.51 \pm 0.00	94.54 \pm 0.00	95.73 \pm 0.00	95.64 \pm 0.00
	rand rand	54.31 \pm 0.00	79.98 \pm 0.00	82.55 \pm 0.00	85.95 \pm 0.00	89.31 \pm 0.00	91.99 \pm 0.00	93.24 \pm 0.00	91.92 \pm 0.00
	randrow randrow	54.24 \pm 0.00	79.67 \pm 0.00	82.21 \pm 0.00	85.25 \pm 0.00	88.46 \pm 0.00	90.58 \pm 0.00	90.48 \pm 0.00	77.27 \pm 0.00
Musr	old new	54.31 \pm 0.00	80.85 \pm 0.00	83.89 \pm 0.00	87.48 \pm 0.00	89.91 \pm 0.00	92.10 \pm 0.00	91.37 \pm 0.00	81.50 \pm 0.00
	small large	54.52 \pm 0.00	78.94 \pm 0.00	80.25 \pm 0.00	81.51 \pm 0.00	83.37 \pm 0.00	84.56 \pm 0.00	83.92 \pm 0.00	73.93 \pm 0.00
	unofficial official	53.93 \pm 0.00	72.36 \pm 0.00	74.15 \pm 0.00	77.92 \pm 0.00	80.33 \pm 0.00	82.58 \pm 0.00	82.75 \pm 0.00	71.44 \pm 0.00
	randcol randcol	54.04 \pm 0.00	79.74 \pm 0.00	82.29 \pm 0.00	85.83 \pm 0.00	89.18 \pm 0.00	91.58 \pm 0.00	92.72 \pm 0.00	91.41 \pm 0.00
	L	54.39 \pm 0.00	79.58 \pm 0.00	82.06 \pm 0.00	85.09 \pm 0.00	87.87 \pm 0.00	90.05 \pm 0.00	89.46 \pm 0.00	78.98 \pm 0.00
	rand rand	54.53 \pm 0.00	87.15 \pm 0.00	88.80 \pm 0.00	90.78 \pm 0.00	92.68 \pm 0.00	94.17 \pm 0.00	94.09 \pm 0.00	90.13 \pm 0.00
Gpqa	randrow randrow	53.85 \pm 0.00	86.83 \pm 0.00	88.43 \pm 0.00	90.25 \pm 0.00	91.67 \pm 0.00	92.21 \pm 0.00	89.21 \pm 0.00	80.56 \pm 0.00
	old new	54.46 \pm 0.00	87.08 \pm 0.00	88.65 \pm 0.00	90.77 \pm 0.00	92.54 \pm 0.00	92.94 \pm 0.00	89.71 \pm 0.00	80.26 \pm 0.00
	small large	53.33 \pm 0.00	86.07 \pm 0.00	87.78 \pm 0.00	88.49 \pm 0.00	89.14 \pm 0.00	89.13 \pm 0.00	85.78 \pm 0.00	77.93 \pm 0.00
	unofficial official	53.47 \pm 0.00	85.15 \pm 0.00	86.15 \pm 0.00	87.26 \pm 0.00	88.85 \pm 0.00	89.11 \pm 0.00	86.00 \pm 0.00	76.69 \pm 0.00
	randcol randcol	55.16 \pm 0.00	86.71 \pm 0.00	88.23 \pm 0.00	90.34 \pm 0.00	92.33 \pm 0.00	93.82 \pm 0.00	94.39 \pm 0.00	93.22 \pm 0.00
	L	53.63 \pm 0.00	87.19 \pm 0.00	88.24 \pm 0.00	90.30 \pm 0.00	91.70 \pm 0.00	91.65 \pm 0.00	90.30 \pm 0.00	83.05 \pm 0.00

D PROMPT FOR INTERPRETING DIMENSION AND CLUSTER

D.1 PROMPT FOR INTERPRETING UNIDIMENSION ITEMS AND THEIR UNDERLYING CONSTRUCTS

(0) Analyze the following positive and negative examples to identify the underlying construct that appear in positive examples and not in negative examples.

POSITIVE EXAMPLES (items that load on this dimension):

(Question 1) question: Mr. Gribbonearns \$10,000 annually as an accountant and pays \$1500 of this amount to the government in taxes. Mr. Doyle earns \$50,000 as a frankfurter vendor, of which he takes home \$45,000

after taxes. Is the tax structure here progressive, proportional, or regressive?

Options: ['Flat rate', 'Graduated', 'Indexed', 'Variable', 'Progressive', 'Regressive', 'Proportional', 'Tiered', 'Inverted', 'None of the above']

(Question 2) question: The power output of a laser is measured in units of watts (W), where one watt is equal to one joule per second. $\$ \leftarrow 1 \cdot \mathbf{W} = 1 \cdot \mathbf{W} \cdot \mathbf{J} \cdot \mathbf{m} \cdot \mathbf{s}^{-1}$. What is

the number of photons emitted per second by a \$1.00 \mathbf{W} nitrogen laser? The wavelength emitted by a nitrogen laser is \$337 \mathbf{nm} .

Options: ['1.70 $\$ \cdot 10^{15}$ \text{ photon}', '1.20 $\$ \cdot 10^{15}$ \text{ photon}', '3.00 $\$ \cdot 10^{15}$ \text{ photon}', '2.50 $\$ \cdot 10^{15}$ \text{ photon}']

(Question 3) question: Suppose that the weights of trucks traveling on the interstate highway system are

normally distributed. If 70% of the trucks weigh more than 12,000 pounds and 80% weigh more than

10,000 pounds, what are the mean and standard deviation for the weights of trucks traveling on the interstate system?

Options: ['= 14,900', '= 6100', '= 16,300', '= 6800', '= 14,500', '= 5900', '= 15,300', '= 6300', '= 15,500', '= 6400', '= 15,100', '= 6200', '= 15,900', '= 6600', '= 14,700', '= 6000', '= 16,100', '= 6700', '= 15,700', '= 6500']

(Question 4) question: One morning, an employee arrived at work and found his boss passed out on the floor of his office. The employee determined that his boss was not breathing, immediately rendered CPR to his boss, and succeeded in reviving him. The employee then called the paramedics, who took the boss to

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

Table 4: Performance of various rank models on Log Probability and MAE

Dataset	Masking	K=0	K=1	K=2	K=4	K=8	K=16	K=32	K=64	K=128	
AUC (x 100)											
HF 1	rand rand	63.57 ± 0.00	87.84 ± 0.00	89.33 ± 0.00	90.75 ± 0.00	92.77 ± 0.00	94.60 ± 0.00	95.98 ± 0.00	96.97 ± 0.00	96.47 ± 0.00	
	randrow randrow	63.76 ± 0.00	87.82 ± 0.00	89.23 ± 0.00	90.75 ± 0.00	92.76 ± 0.00	94.56 ± 0.00	95.93 ± 0.00	95.71 ± 0.00	83.55 ± 0.00	
	old new	63.41 ± 0.00	86.96 ± 0.00	89.30 ± 0.00	90.61 ± 0.00	93.32 ± 0.00	94.51 ± 0.00	96.01 ± 0.00	96.20 ± 0.00	84.22 ± 0.00	
	small large	59.99 ± 0.00	84.11 ± 0.00	85.38 ± 0.00	84.95 ± 0.00	86.04 ± 0.00	88.26 ± 0.00	90.53 ± 0.00	90.40 ± 0.00	96.20 ± 0.00	
	unofficial official	54.86 ± 0.00	84.91 ± 0.00	85.68 ± 0.00	86.28 ± 0.00	87.82 ± 0.00	89.89 ± 0.00	91.68 ± 0.00	92.09 ± 0.00	79.45 ± 0.00	
	randcol randcol	63.77 ± 0.00	87.85 ± 0.00	89.38 ± 0.00	90.82 ± 0.00	92.81 ± 0.00	94.61 ± 0.00	95.98 ± 0.00	96.94 ± 0.00	96.69 ± 0.00	
Mmlu Pro	mmlu bbb else	58.48 ± 0.00	83.93 ± 0.00	85.00 ± 0.00	85.84 ± 0.00	87.22 ± 0.00	88.90 ± 0.00	90.32 ± 0.00	91.41 ± 0.00	90.37 ± 0.00	
	L	64.05 ± 0.00	88.05 ± 0.00	89.59 ± 0.00	90.59 ± 0.00	92.82 ± 0.00	94.55 ± 0.00	95.99 ± 0.00	96.54 ± 0.00	83.33 ± 0.00	
	rand rand	66.10 ± 0.00	89.00 ± 0.00	90.58 ± 0.00	92.01 ± 0.00	93.86 ± 0.00	95.50 ± 0.00	96.62 ± 0.00	97.39 ± 0.00	96.86 ± 0.00	
	randrow randrow	66.31 ± 0.00	88.99 ± 0.00	90.48 ± 0.00	91.91 ± 0.00	93.84 ± 0.00	95.46 ± 0.00	96.59 ± 0.00	96.18 ± 0.00	83.88 ± 0.00	
	old new	67.83 ± 0.00	89.55 ± 0.00	91.29 ± 0.00	92.37 ± 0.00	94.66 ± 0.00	96.13 ± 0.00	96.98 ± 0.00	97.39 ± 0.00	84.79 ± 0.00	
	small large	61.28 ± 0.00	84.94 ± 0.00	86.55 ± 0.00	86.28 ± 0.00	88.07 ± 0.00	89.30 ± 0.00	90.01 ± 0.00	90.64 ± 0.00	79.00 ± 0.00	
Openlm Math	unofficial official	68.51 ± 0.00	85.78 ± 0.00	86.72 ± 0.00	87.37 ± 0.00	89.09 ± 0.00	90.93 ± 0.00	92.43 ± 0.00	92.44 ± 0.00	78.51 ± 0.00	
	randcol randcol	66.32 ± 0.00	88.98 ± 0.00	90.60 ± 0.00	91.99 ± 0.00	93.90 ± 0.00	95.53 ± 0.00	96.62 ± 0.00	97.39 ± 0.00	97.08 ± 0.00	
	mmlu bbb else	—	—	—	—	—	—	—	—	—	
	L	66.41 ± 0.00	89.23 ± 0.00	90.82 ± 0.00	92.25 ± 0.00	93.90 ± 0.00	95.46 ± 0.00	96.55 ± 0.00	96.99 ± 0.00	83.59 ± 0.00	
	rand rand	81.04 ± 0.00	86.83 ± 0.00	87.46 ± 0.00	87.81 ± 0.00	89.19 ± 0.00	90.68 ± 0.00	92.38 ± 0.00	92.60 ± 0.00	90.05 ± 0.00	
	randrow randrow	81.45 ± 0.00	87.58 ± 0.00	88.13 ± 0.00	89.10 ± 0.00	89.82 ± 0.00	91.32 ± 0.00	92.47 ± 0.00	92.47 ± 0.00	66.39 ± 0.00	
Musr	old new	79.98 ± 0.00	87.15 ± 0.00	87.83 ± 0.00	88.67 ± 0.00	89.12 ± 0.00	89.83 ± 0.00	91.93 ± 0.00	91.53 ± 0.00	70.10 ± 0.00	
	small large	75.91 ± 0.00	80.14 ± 0.00	81.82 ± 0.00	82.33 ± 0.00	82.12 ± 0.00	82.96 ± 0.00	85.81 ± 0.00	85.55 ± 0.00	70.64 ± 0.00	
	unofficial official	84.06 ± 0.00	85.87 ± 0.00	85.92 ± 0.00	86.24 ± 0.00	87.26 ± 0.00	88.85 ± 0.00	90.46 ± 0.00	90.39 ± 0.00	56.59 ± 0.00	
	randcol randcol	81.33 ± 0.00	86.16 ± 0.00	86.83 ± 0.00	87.71 ± 0.00	88.68 ± 0.00	89.28 ± 0.00	91.84 ± 0.00	92.19 ± 0.00	90.11 ± 0.00	
	mmlu bbb else	81.23 ± 0.00	86.50 ± 0.00	87.09 ± 0.00	87.55 ± 0.00	88.42 ± 0.00	89.09 ± 0.00	89.58 ± 0.00	89.77 ± 0.00	86.76 ± 0.00	
	L	80.34 ± 0.00	87.12 ± 0.00	87.62 ± 0.00	88.25 ± 0.00	89.47 ± 0.00	90.92 ± 0.00	92.18 ± 0.00	92.91 ± 0.00	65.12 ± 0.00	
Bbth	rand rand	54.43 ± 0.00	86.39 ± 0.00	87.88 ± 0.00	88.85 ± 0.00	90.53 ± 0.00	91.90 ± 0.00	93.31 ± 0.00	94.68 ± 0.00	94.64 ± 0.00	
	randrow randrow	54.50 ± 0.00	86.40 ± 0.00	87.80 ± 0.00	88.74 ± 0.00	90.57 ± 0.00	91.89 ± 0.00	93.20 ± 0.00	92.89 ± 0.00	80.96 ± 0.00	
	old new	54.87 ± 0.00	86.48 ± 0.00	88.34 ± 0.00	88.78 ± 0.00	90.68 ± 0.00	92.19 ± 0.00	93.12 ± 0.00	93.91 ± 0.00	81.02 ± 0.00	
	small large	53.84 ± 0.00	85.44 ± 0.00	86.36 ± 0.00	86.28 ± 0.00	86.49 ± 0.00	87.46 ± 0.00	87.77 ± 0.00	88.31 ± 0.00	78.74 ± 0.00	
	unofficial official	53.79 ± 0.00	83.75 ± 0.00	84.82 ± 0.00	84.96 ± 0.00	85.83 ± 0.00	86.59 ± 0.00	87.57 ± 0.00	87.71 ± 0.00	78.10 ± 0.00	
	randcol randcol	55.36 ± 0.00	85.25 ± 0.00	86.90 ± 0.00	87.84 ± 0.00	89.90 ± 0.00	91.24 ± 0.00	92.80 ± 0.00	94.17 ± 0.00	93.13 ± 0.00	
Bbth	mmlu bbb else	54.27 ± 0.00	86.47 ± 0.00	87.90 ± 0.00	88.80 ± 0.00	90.02 ± 0.00	91.38 ± 0.00	92.76 ± 0.00	93.81 ± 0.00	93.13 ± 0.00	
	L	54.57 ± 0.00	87.44 ± 0.00	88.93 ± 0.00	89.64 ± 0.00	91.21 ± 0.00	92.30 ± 0.00	93.57 ± 0.00	94.31 ± 0.00	81.80 ± 0.00	
	rand rand	62.59 ± 0.00	85.64 ± 0.00	87.32 ± 0.00	89.11 ± 0.00	91.46 ± 0.00	94.05 ± 0.00	95.83 ± 0.00	97.11 ± 0.00	96.64 ± 0.00	
	randrow randrow	62.71 ± 0.00	85.58 ± 0.00	87.19 ± 0.00	91.48 ± 0.00	94.00 ± 0.00	95.83 ± 0.00	95.79 ± 0.00	96.05 ± 0.00	83.70 ± 0.00	
	old new	63.78 ± 0.00	86.19 ± 0.00	88.00 ± 0.00	89.43 ± 0.00	92.14 ± 0.00	94.58 ± 0.00	96.03 ± 0.00	96.95 ± 0.00	84.42 ± 0.00	
	small large	59.98 ± 0.00	82.31 ± 0.00	82.90 ± 0.00	82.08 ± 0.00	84.22 ± 0.00	87.16 ± 0.00	89.45 ± 0.00	90.65 ± 0.00	78.65 ± 0.00	
Ifeval	unofficial official	63.35 ± 0.00	81.87 ± 0.00	82.53 ± 0.00	83.53 ± 0.00	85.20 ± 0.00	85.20 ± 0.00	88.69 ± 0.00	91.27 ± 0.00	92.38 ± 0.00	80.67 ± 0.00
	randcol randcol	62.83 ± 0.00	85.78 ± 0.00	87.50 ± 0.00	89.50 ± 0.00	91.54 ± 0.00	94.02 ± 0.00	95.83 ± 0.00	97.04 ± 0.00	96.86 ± 0.00	
	mmlu bbb else	—	—	—	—	—	—	—	—	—	
	L	63.27 ± 0.00	85.96 ± 0.00	87.72 ± 0.00	89.30 ± 0.00	91.49 ± 0.00	93.90 ± 0.00	95.72 ± 0.00	96.59 ± 0.00	83.39 ± 0.00	
	rand rand	74.71 ± 0.00	82.06 ± 0.00	82.71 ± 0.00	87.68 ± 0.00	89.13 ± 0.00	91.46 ± 0.00	94.05 ± 0.00	92.43 ± 0.00	91.73 ± 0.00	
	randrow randrow	74.62 ± 0.00	81.96 ± 0.00	82.50 ± 0.00	83.48 ± 0.00	87.49 ± 0.00	88.87 ± 0.00	91.27 ± 0.00	91.60 ± 0.00	79.36 ± 0.00	
Gpq	old new	76.59 ± 0.00	81.61 ± 0.00	82.05 ± 0.00	82.52 ± 0.00	86.52 ± 0.00	88.47 ± 0.00	91.04 ± 0.00	92.00 ± 0.00	80.80 ± 0.00	
	small large	76.40 ± 0.00	79.67 ± 0.00	79.70 ± 0.00	79.70 ± 0.00	80.20 ± 0.00	81.30 ± 0.00	87.00 ± 0.00	87.70 ± 0.00	77.95 ± 0.00	
	unofficial official	75.12 ± 0.00	81.79 ± 0.00	81.96 ± 0.00	81.72 ± 0.00	83.83 ± 0.00	84.98 ± 0.00	88.30 ± 0.00	89.22 ± 0.00	87.00 ± 0.00	
	randcol randcol	74.80 ± 0.00	82.49 ± 0.00	83.06 ± 0.00	83.85 ± 0.00	88.13 ± 0.00	89.28 ± 0.00	91.86 ± 0.00	92.29 ± 0.00	92.50 ± 0.00	
	mmlu bbb else	74.78 ± 0.00	81.83 ± 0.00	82.48 ± 0.00	83.06 ± 0.00	84.29 ± 0.00	86.94 ± 0.00	88.45 ± 0.00	89.19 ± 0.00	87.87 ± 0.00	
	L	74.21 ± 0.00	83.21 ± 0.00	84.03 ± 0.00	85.56 ± 0.00	88.39 ± 0.00	91.46 ± 0.00	91.74 ± 0.00	92.49 ± 0.00	80.82 ± 0.00	
Ifeval	rand rand	54.43 ± 0.00	84.43 ± 0.00	75.39 ± 0.00	77.44 ± 0.00	78.98 ± 0.00	83.99 ± 0.00	86.36 ± 0.00	89.48 ± 0.00	93.01 ± 0.00	
	randrow randrow	54.50 ± 0.00	75.00 ± 0.00	76.88 ± 0.00	78.00 ± 0.00	83.50 ± 0.00	86.36 ± 0.00	89.40 ± 0.00	90.90 ± 0.00	92.30 ± 0.00	
	old new	54.91 ± 0.00	76.53 ± 0.00	78.35 ± 0.00	79.13 ± 0.00	83.44 ± 0.00	87.31 ± 0.00	90.00 ± 0.00	92.32 ± 0.00	75.59 ± 0.00	
	small large	54.88 ± 0.00	73.74 ± 0.00	75.49 ± 0.00	74.54 ± 0.00	76.88 ± 0.00	77.99 ± 0.00	82.50 ± 0.00	83.80 ± 0.00	70.21 ± 0.00	
	unofficial official	54.66 ± 0.00	69.77 ± 0.00	70.08 ± 0.00	70.26 ± 0.00	72.77 ± 0.00	74.75 ± 0.00	77.80 ± 0.00	81.65 ± 0.00	68.26 ± 0.00	
	randcol randcol	50.56 ± 0.00	74.59 ± 0.00	76.25 ± 0.00	78.01 ± 0.00	83.34 ± 0.00	85.77 ± 0.00	89.17 ± 0.00	92.57 ± 0.00	92.94 ± 0.00	
Gpq	mmlu bbb else	54.42 ± 0.00	75.64 ± 0.00	77.34 ± 0.00	78.88 ± 0.00	81.23 ± 0.00	83.97 ± 0.00	86.36 ± 0.00	88.46 ± 0.00	87.98 ± 0.00	
	L	55.12 ± 0.00	74.91 ± 0.00	76.72 ± 0.00	78.21 ± 0.00	83.74 ± 0.00	85.87 ± 0.00	88.87 ± 0.00	91.68 ± 0.00	73.05 ± 0.00	
	rand rand	80.93 ± 0.00	92.03 ± 0.00	92.51 ± 0.00	93.07 ± 0.00	93.97 ± 0.00	95.42 ± 0.00	97.62 ± 0.00	97.77 ± 0.00	98.32 ± 0.00	
	randrow randrow	81.33 ± 0.00	91.96 ± 0.00	92.42 ± 0.00	93.00 ± 0.00	93.45 ± 0.00	94.63 ± 0.00	96.70 ± 0.00	97.06 ± 0.00	99.38 ± 0.00	
	old new	79.76 ± 0.00</									

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824

Table 6: Performance of various rank models on Log Probability and MAE

Dataset	Masking	K=0	K=1	K=2	K=4	K=8	K=16	K=32	K=64	K=128
AUC (x 100)										
HF 1	rand rand	63.57 ± 0.00	87.87 ± 0.00	89.38 ± 0.00	90.82 ± 0.00	92.89 ± 0.00	94.81 ± 0.00	96.35 ± 0.00	97.74 ± 0.00	98.88 ± 0.00
	randrow randrow	63.76 ± 0.00	87.87 ± 0.00	89.40 ± 0.00	90.86 ± 0.00	92.88 ± 0.00	94.82 ± 0.00	96.39 ± 0.00	97.74 ± 0.00	98.85 ± 0.00
	old new	65.1 ± 0.00	87.75 ± 0.00	89.20 ± 0.00	90.71 ± 0.00	92.61 ± 0.00	94.60 ± 0.00	96.24 ± 0.00	97.63 ± 0.00	98.89 ± 0.00
	small large	59.98 ± 0.00	87.47 ± 0.00	89.14 ± 0.00	90.79 ± 0.00	92.79 ± 0.00	94.86 ± 0.00	96.51 ± 0.00	97.84 ± 0.00	99.04 ± 0.00
	unofficial official	64.86 ± 0.00	88.11 ± 0.00	89.71 ± 0.00	91.24 ± 0.00	93.29 ± 0.00	95.21 ± 0.00	96.72 ± 0.00	98.00 ± 0.00	99.04 ± 0.00
	randcol randcol	63.77 ± 0.00	87.86 ± 0.00	89.36 ± 0.00	90.84 ± 0.00	92.87 ± 0.00	94.79 ± 0.00	96.5 ± 0.00	97.70 ± 0.00	98.79 ± 0.00
	mmlu bbb else	58.48 ± 0.00	88.41 ± 0.00	89.97 ± 0.00	91.53 ± 0.00	93.52 ± 0.00	95.42 ± 0.00	96.90 ± 0.00	98.10 ± 0.00	99.08 ± 0.00
Mmlu Pro	L	64.05 ± 0.00	87.82 ± 0.00	89.32 ± 0.00	90.83 ± 0.00	92.87 ± 0.00	94.82 ± 0.00	96.38 ± 0.00	97.74 ± 0.00	98.91 ± 0.00
	rand rand	66.10 ± 0.00	89.00 ± 0.00	90.58 ± 0.00	92.01 ± 0.00	93.86 ± 0.00	95.50 ± 0.00	96.62 ± 0.00	97.39 ± 0.00	96.86 ± 0.00
	randrow randrow	66.31 ± 0.00	88.99 ± 0.00	90.48 ± 0.00	91.91 ± 0.00	93.84 ± 0.00	95.46 ± 0.00	96.58 ± 0.00	96.18 ± 0.00	83.88 ± 0.00
	old new	67.83 ± 0.00	89.55 ± 0.00	91.29 ± 0.00	92.37 ± 0.00	94.66 ± 0.00	96.13 ± 0.00	96.96 ± 0.00	97.39 ± 0.00	84.79 ± 0.00
	small large	61.28 ± 0.00	84.94 ± 0.00	86.55 ± 0.00	88.38 ± 0.00	88.07 ± 0.00	89.30 ± 0.00	90.01 ± 0.00	90.64 ± 0.00	79.00 ± 0.00
	unofficial official	68.51 ± 0.00	85.78 ± 0.00	86.72 ± 0.00	87.37 ± 0.00	89.09 ± 0.00	90.93 ± 0.00	92.43 ± 0.00	92.44 ± 0.00	78.51 ± 0.00
	randcol randcol	66.32 ± 0.00	88.98 ± 0.00	90.60 ± 0.00	91.99 ± 0.00	93.90 ± 0.00	95.53 ± 0.00	96.62 ± 0.00	97.39 ± 0.00	79.08 ± 0.00
Openlim Math	mmlu bbb else	66.41 ± 0.00	89.23 ± 0.00	90.82 ± 0.00	92.25 ± 0.00	93.90 ± 0.00	95.46 ± 0.00	96.55 ± 0.00	96.99 ± 0.00	83.59 ± 0.00
	L	81.04 ± 0.00	86.83 ± 0.00	87.46 ± 0.00	87.81 ± 0.00	89.19 ± 0.00	90.58 ± 0.00	92.38 ± 0.00	92.66 ± 0.00	90.85 ± 0.00
	rand rand	81.45 ± 0.00	87.83 ± 0.00	88.13 ± 0.00	88.52 ± 0.00	89.82 ± 0.00	91.32 ± 0.00	92.47 ± 0.00	96.39 ± 0.00	96.39 ± 0.00
	randrow randrow	79.93 ± 0.00	87.15 ± 0.00	87.83 ± 0.00	88.67 ± 0.00	89.12 ± 0.00	89.83 ± 0.00	91.91 ± 0.00	91.53 ± 0.00	70.10 ± 0.00
	old new	75.91 ± 0.00	80.14 ± 0.00	81.82 ± 0.00	82.32 ± 0.00	82.12 ± 0.00	82.96 ± 0.00	85.81 ± 0.00	85.55 ± 0.00	70.64 ± 0.00
	small large	84.06 ± 0.00	85.87 ± 0.00	85.92 ± 0.00	86.24 ± 0.00	87.26 ± 0.00	88.85 ± 0.00	90.46 ± 0.00	90.39 ± 0.00	56.59 ± 0.00
	unofficial official	81.33 ± 0.00	86.16 ± 0.00	86.83 ± 0.00	87.71 ± 0.00	88.68 ± 0.00	89.28 ± 0.00	91.84 ± 0.00	92.19 ± 0.00	90.11 ± 0.00
Musr	randcol randcol	81.23 ± 0.00	86.50 ± 0.00	87.09 ± 0.00	87.55 ± 0.00	88.42 ± 0.00	89.09 ± 0.00	89.58 ± 0.00	89.77 ± 0.00	86.76 ± 0.00
	mmlu bbb else	80.34 ± 0.00	87.12 ± 0.00	87.62 ± 0.00	88.25 ± 0.00	89.47 ± 0.00	90.92 ± 0.00	92.18 ± 0.00	92.91 ± 0.00	65.12 ± 0.00
	L	54.43 ± 0.00	86.39 ± 0.00	87.88 ± 0.00	88.85 ± 0.00	90.53 ± 0.00	91.90 ± 0.00	93.31 ± 0.00	94.68 ± 0.00	94.64 ± 0.00
	rand rand	54.50 ± 0.00	86.40 ± 0.00	87.80 ± 0.00	88.74 ± 0.00	90.57 ± 0.00	91.89 ± 0.00	93.20 ± 0.00	92.89 ± 0.00	80.96 ± 0.00
	randrow randrow	54.87 ± 0.00	86.48 ± 0.00	88.34 ± 0.00	88.78 ± 0.00	90.68 ± 0.00	92.19 ± 0.00	93.12 ± 0.00	93.91 ± 0.00	81.02 ± 0.00
	old new	53.84 ± 0.00	85.44 ± 0.00	86.36 ± 0.00	86.26 ± 0.00	86.49 ± 0.00	87.46 ± 0.00	87.77 ± 0.00	88.31 ± 0.00	78.74 ± 0.00
	small large	53.79 ± 0.00	83.75 ± 0.00	84.82 ± 0.00	84.96 ± 0.00	85.83 ± 0.00	86.59 ± 0.00	87.57 ± 0.00	87.71 ± 0.00	78.10 ± 0.00
Bbb	unofficial official	55.36 ± 0.00	85.25 ± 0.00	86.90 ± 0.00	87.84 ± 0.00	89.90 ± 0.00	91.24 ± 0.00	92.80 ± 0.00	94.17 ± 0.00	94.47 ± 0.00
	randcol randcol	54.27 ± 0.00	86.47 ± 0.00	87.90 ± 0.00	88.89 ± 0.00	90.02 ± 0.00	91.38 ± 0.00	92.76 ± 0.00	93.81 ± 0.00	93.13 ± 0.00
	mmlu bbb else	54.57 ± 0.00	87.44 ± 0.00	88.93 ± 0.00	89.64 ± 0.00	91.21 ± 0.00	92.30 ± 0.00	93.57 ± 0.00	94.31 ± 0.00	81.80 ± 0.00
	L	62.59 ± 0.00	85.64 ± 0.00	87.32 ± 0.00	89.11 ± 0.00	91.46 ± 0.00	94.05 ± 0.00	95.83 ± 0.00	97.11 ± 0.00	96.64 ± 0.00
	rand rand	62.71 ± 0.00	85.58 ± 0.00	87.19 ± 0.00	91.48 ± 0.00	94.00 ± 0.00	95.83 ± 0.00	95.79 ± 0.00	95.70 ± 0.00	93.70 ± 0.00
	randrow randrow	63.78 ± 0.00	86.19 ± 0.00	88.00 ± 0.00	89.43 ± 0.00	92.14 ± 0.00	94.58 ± 0.00	96.03 ± 0.00	96.95 ± 0.00	84.42 ± 0.00
	old new	59.98 ± 0.00	82.31 ± 0.00	82.90 ± 0.00	82.08 ± 0.00	84.22 ± 0.00	87.16 ± 0.00	89.45 ± 0.00	90.65 ± 0.00	78.65 ± 0.00
Ifeval	small large	74.40 ± 0.00	79.55 ± 0.00	79.17 ± 0.00	77.94 ± 0.00	79.63 ± 0.00	81.14 ± 0.00	83.77 ± 0.00	87.10 ± 0.00	77.95 ± 0.00
	unofficial official	75.12 ± 0.00	81.79 ± 0.00	81.96 ± 0.00	81.72 ± 0.00	83.83 ± 0.00	84.98 ± 0.00	88.36 ± 0.00	89.22 ± 0.00	79.44 ± 0.00
	randcol randcol	74.80 ± 0.00	82.49 ± 0.00	83.06 ± 0.00	83.85 ± 0.00	88.13 ± 0.00	89.28 ± 0.00	91.86 ± 0.00	92.29 ± 0.00	92.50 ± 0.00
	mmlu bbb else	74.78 ± 0.00	81.83 ± 0.00	82.48 ± 0.00	83.06 ± 0.00	84.29 ± 0.00	86.94 ± 0.00	88.45 ± 0.00	89.19 ± 0.00	87.87 ± 0.00
	L	74.21 ± 0.00	83.21 ± 0.00	84.03 ± 0.00	85.56 ± 0.00	88.39 ± 0.00	89.46 ± 0.00	91.74 ± 0.00	92.49 ± 0.00	80.82 ± 0.00
	rand rand	54.43 ± 0.00	75.39 ± 0.00	77.21 ± 0.00	78.99 ± 0.00	83.99 ± 0.00	86.36 ± 0.00	89.48 ± 0.00	92.80 ± 0.00	93.01 ± 0.00
	randrow randrow	54.56 ± 0.00	75.23 ± 0.00	76.88 ± 0.00	78.45 ± 0.00	83.85 ± 0.00	86.17 ± 0.00	88.95 ± 0.00	90.62 ± 0.00	74.26 ± 0.00
Gpqa	old new	54.91 ± 0.00	76.53 ± 0.00	78.35 ± 0.00	79.13 ± 0.00	83.44 ± 0.00	87.31 ± 0.00	90.04 ± 0.00	92.32 ± 0.00	74.59 ± 0.00
	small large	54.88 ± 0.00	73.74 ± 0.00	75.49 ± 0.00	74.54 ± 0.00	76.88 ± 0.00	77.99 ± 0.00	82.50 ± 0.00	83.80 ± 0.00	70.21 ± 0.00
	unofficial official	54.35 ± 0.00	69.47 ± 0.00	70.81 ± 0.00	78.01 ± 0.00	83.34 ± 0.00	85.75 ± 0.00	89.17 ± 0.00	92.57 ± 0.00	69.00 ± 0.00
	randcol randcol	53.35 ± 0.00	74.59 ± 0.00	76.25 ± 0.00	78.01 ± 0.00	83.34 ± 0.00	85.75 ± 0.00	89.17 ± 0.00	92.94 ± 0.00	69.00 ± 0.00
	mmlu bbb else	54.42 ± 0.00	75.64 ± 0.00	77.34 ± 0.00	78.86 ± 0.00	81.23 ± 0.00	83.97 ± 0.00	86.36 ± 0.00	88.46 ± 0.00	87.98 ± 0.00
	L	55.12 ± 0.00	74.91 ± 0.00	76.72 ± 0.00	78.21 ± 0.00	83.74 ± 0.00	85.87 ± 0.00	88.87 ± 0.00	91.68 ± 0.00	73.05 ± 0.00

851
852
853
854
855
856
857
858
859
860
861
862
863

864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917

→ the hospital. Two week later, the boss returned to work. Grateful for the employee's quick action in
 → saving his life, the boss said to the employee, "You'll have a job with me for life. "Which of the
 → following best characterizes the legal relationship between the boss and the employee on that date?
 Options: ["Since the employee gratuitously rendered assistance to the boss, there was insufficient
 → consideration to support the boss's subsequent promise to provide the employee with lifetime
 → employment.", "The boss had a duty to pay the employee reasonable compensation for saving his life,
 → based upon a contract implied-in-fact.", "The employee had a valid enforceable contract due to the
 → boss's promise.", "The employee's act of saving the boss's life constituted a legally enforceable
 → contract.", "The employee's act of saving the boss's life was sufficient past consideration to
 → render enforceable the boss's subsequent promise.", "The boss's promise was not legally binding as
 → it was made under duress.", "The employee's act of saving the boss's life was a gift and did not
 → create a contractual obligation.", "The boss's promise was voidable due to lack of consideration.",
 → "As per the boss's oral promise, the employee had an enforceable lifetime employment contract.",
 → "The boss was under no legal obligation to promise the employee anything."]
(Question 5) question: A spring of force constant k is stretched a certain distance. It takes twice as much
 → work to stretch a second spring by half this distance. The force constant of the second spring is
 Options: ['8k', '3k', 'k/4', 'k', '4k', '2k', 'k/2', '5k', '16k']
(Question 6) question: A plate at 60 °C, is cooled by forced convection from an airstream. The steady flow
 → velocity of the air stream is observed to be 2m/sec with the temperature and pressure as 27 °C and 1
 → atm., respectively. Find the heat transfer per unit width from the leading edge to a distance of (a)
 → 20 cm, and (b) 40 cm. Also find (c) the drag force over the 40 cm distance.
 Options: ['75 W, 105 W, 0.004 N', '81.19 W, 114.8 W, 0.0055 N', '85 W, 120 W, 0.005 N', '90 W, 130 W, 0.007 N
 → ', '82 W, 115 W, 0.005 N', '83 W, 118 W, 0.0065 N', '86 W, 125 W, 0.0058 N', '78 W, 112 W, 0.0045 N
 → ', '80 W, 110 W, 0.006 N', '79.5 W, 108 W, 0.0075 N']
(Question 7) question: The 'new racism' refers to:
 Options: ['a subtler form of prejudice, masked by nationalist pride', 'a trend of neo-nazism masked as
 → patriotism', 'a form of racism that only exists within the realm of social media', 'an anti-fascist
 → movement which challenges nationalist politics', 'an increased awareness and sensitivity towards
 → racism in contemporary society', 'a post-modern deconstruction of racist ideas to reveal their lack
 → of depth', 'a radical form of racism that encourages open discrimination', 'a form of racism
 → practiced by newly established countries', 'racist practices found in newly emerging areas of social
 → life, such as cyberspace', 'the rise of racism in new political parties']
(Question 8) question: Which of the following statements is correct regarding the Federal Unemployment Tax Act
 → ?
 Options: ['The unemployment insurance system is administered by the states through their employment laws.',
 → 'The Act is intended to provide financial assistance to unemployed workers who have been laid off due
 → to seasonal changes in demand.', 'An employee who resigns regardless of cause is eligible for
 → unemployment benefits.', 'An employee who is fired for cause is eligible for unemployment benefits
 → .', 'The Act is intended to assist workers who are permanently out of work and need assistance in
 → supporting themselves.', 'The Act mandates that employers provide health insurance to employees who
 → are unemployed.', 'The federal unemployment system is funded solely by employee taxes.', 'The
 → federal unemployment system is funded by both employer and employee taxes.', 'The Act also covers
 → compensation for self-employed individuals.', 'The Act is enforced by the Federal Labor Relations
 → Authority.'][br/>
(Question 9) question: A sample of polymer contains 0.50 mole fraction with molecular weight 100,000 and 0.50
 → mole fraction with molecular weight 200,000. Calculate (a) the number average molecular weight, M_n
 → and (b) the weight average molecular weight, M_w .
 Options: [' $M_n = 200,000$ g/mole, $M_w = 1.0 \times 10^5$ g/mole', ' $M_n = 150,000$ g/mole, $M_w = 1.67 \times 10^5$ g/mole
 → ', ' $M_n = 250,000$ g/mole, $M_w = 0.67 \times 10^5$ g/mole', ' $M_n = 150,000$ g/mole, $M_w = 2.0 \times 10^5$ g/
 → mole', ' $M_n = 150,000$ g/mole, $M_w = 1.5 \times 10^5$ g/mole', ' $M_n = 100,000$ g/mole, $M_w = 2.0 \times 10^5$ g/
 → mole', ' $M_n = 125,000$ g/mole, $M_w = 1.75 \times 10^5$ g/mole', ' $M_n = 200,000$ g/mole, $M_w = 1.75 \times 10^5$
 → g/mole', ' $M_n = 175,000$ g/mole, $M_w = 1.25 \times 10^5$ g/mole', ' $M_n = 100,000$ g/mole, $M_w = 1.5 \times 10^5$
 → g/mole']
(Question 10) question: How do you describe a work group that comprises workers with demarcated tasks and
 → distant line managers?
 Options: ['Interdependent', 'Co-dependent', 'Cohesive', 'Integrated', 'Synergistic', 'Independent',
 → 'Collaborative', 'Hierarchical', 'Dependent']
NEGATIVE EXAMPLES (items that do NOT load on this dimension):
(Question 1) question: A 360 ° journal bearing 3 in. long, carries a 4 in. diameter shaft with a radial clearance
 → of 0.0025 in. The shaft supports a radial load of 1000 lbs. and at a speed of 500 rpm. The
 → operating temperature for the bearing is 140 F and SAE 20 oil is used for bearing lubrication. Evaluate
 → the following by Petroff's equation: (1) Friction coefficient for the bearing (2) Heat energy lost
 → due to friction in the bearing.
 Options: ['Friction coefficient for the bearing is 0.003284, Heat energy lost due to friction in the bearing
 → is 0.066 hp', 'Friction coefficient for the bearing is 0.005124, Heat energy lost due to friction in
 → the bearing is 0.070 hp', 'Friction coefficient for the bearing is 0.005784, Heat energy lost due
 → to friction in the bearing is 0.065 hp', 'Friction coefficient for the bearing is 0.005234, Heat
 → energy lost due to friction in the bearing is 0.072 hp', 'Friction coefficient for the bearing is
 → 0.002754, Heat energy lost due to friction in the bearing is 0.060 hp', 'Friction coefficient for
 → the bearing is 0.006234, Heat energy lost due to friction in the bearing is 0.075 hp', 'Friction
 → coefficient for the bearing is 0.003964, Heat energy lost due to friction in the bearing is 0.073 hp
 → ', 'Friction coefficient for the bearing is 0.007314, Heat energy lost due to friction in the
 → bearing is 0.082 hp', 'Friction coefficient for the bearing is 0.004564, Heat energy lost due to
 → friction in the bearing is 0.071 hp', 'Friction coefficient for the bearing is 0.004264, Heat energy
 → lost due to friction in the bearing is 0.068 hp']
(Question 2) question: One mole of ideal gas at 300 K is expanded adiabatically and reversibly from 20 to 1
 → atm. What is the final temperature of the gas, assuming $C_V = (3/2) R$ per mole?
 Options: ['135 K', '150 K', '175 K', '110 K', '180 K', '90.5 K', '60 K', '120 K', '165 K', '200 K']
(Question 3) question: The current through an inductor with inductance $L = 10^{-3}$ henry is given as $i_L(t) = 0.1$
 → $\sin 10^6 t$. Find the voltage $V_L(t)$ across this inductor.
 Options: ['0.1cos10^6t', '100cos10^6t', '1cos10^6t', '10^-3cos10^6t', '0.01sin10^6t', '10^4sin10^6t', '10^6
 → cos10^6t', '10sin10^6t', '0.1sin10^6t', '100sin10^6t']
(Question 4) question: The Shang Dynasty laid the foundation for a coherent development of Chinese
 → civilization that lasted well into the:
 Options: ['17th century.', '14th century.', '18th century.', '21st century.', '5th century.', '13th century
 → .', '20th century.', '8th century.', '15th century.', '10th century.'][br/>
(Question 5) question: One message that Nathanson argues we can send by abolishing the death penalty is that
 Options: ['we believe in the possibility of redemption and change.', 'there are some penalties worse than
 → death.', 'we are unable to administer justice effectively.', 'there is no punishment that is
 → proportional to murder.', 'there is always a possibility of judicial errors.', 'we are lenient']

918 → towards heinous crimes.', 'we should prioritize rehabilitation over punishment.', 'none of the above
 919 → ', 'we respect the dignity of human life.', 'we have a flawed criminal justice system.]
 920 (Question 6) question: 2 moles of a monoatomic ideal gas at 45 C are compressed adiabatically and reversibly
 921 → from 75 liters to 20 liters. Calculate q, W, ΔH and ΔE for the process, excluding vi-
 922 → brational contributions.
 923 Options: ['q=0, E=15.00kJ, H=22.50kJ', 'q=11.21kJ, E=0, H=18.68kJ', 'q=0, E=18.68kJ, H=11.21kJ',
 924 → 'q=18.68kJ, E=18.68kJ, H=18.68kJ', 'q=0, E=11.21kJ, H=18.68kJ', 'q=11.21kJ, E=18.68kJ, H
 925 → =0', 'q=22.50kJ, E=15.00kJ, H=0', 'q=0, E=22.50kJ, H=15.00kJ', 'q=18.68kJ, E=11.21kJ, H
 926 → =0', 'q=18.68kJ, E=0, H=11.21kJ']
 927 (Question 7) question: The Shang Dynasty laid the foundation for a coherent development of Chinese
 928 → civilization that lasted well into the:
 929 Options: ['17th century.', '14th century.', '18th century.', '21st century.', '5th century.', '13th century
 930 → .', '20th century.', '8th century.', '15th century.', '10th century.'][br/>
 931 (Question 8) question: For a given level of government spending the federal government usually experiences a
 932 → budget ____ during economic ____ and a budget ____ during economic ____.
 933 Options: ['surplus expansion surplus expansion', 'deficit recession deficit expansion', 'surplus expansion
 934 → deficit recession', 'deficit expansion deficit recession', 'surplus expansion surplus recession',
 935 → deficit recession surplus expansion', 'surplus recession deficit expansion', 'surplus recession
 936 → surplus expansion', 'deficit expansion surplus recession', 'surplus recession deficit recession'][br/>
 937 (Question 9) question: If Young's Modulus for steel is $19 \times 10^11 \text{ dynes/cm}^2$, how much force will be required
 938 → to stretch a sample of wire 1 sq mm in cross section by 0.2% of its original length?
 939 Options: ['22 10^8 dynes', '14 10^8 dynes', '23 10^8 dynes', '21 10^8 dynes', '16 10^8 dynes',
 940 → '19 10^8 dynes', '18 10^8 dynes', '20 10^8 dynes', '17 10^8 dynes', '15 10^8 dynes'][br/>
 941 (Question 10) question: Mr. John Kirk borrowed \$3,000 to pay his bills. He was charged an interest rate of
 942 → $10\% / 2$ % for a 180 day period. How much interest will Mr. Kirk have to pay?
 943 Options: ['\$210.00', '\$105.00', '\$150.00', '\$225.00', '\$120.00', '\$262.50', '\$75.00', '\$315.00', '\$157.50',
 944 → '\$180.00'][br/>
 945 (Question 11) question: A 360 journal bearing 3 in.long,carries a 4 in. diameter shaft with a radial
 946 → clearance of 0.0025 in. The shaft supports a radial load of 1000 lbs. and at a speed of 500 rpm. The
 947 → operatingtemperature for the bearing is 140 F and SAE 20 oilis used for bearing lubrication. Evaluate
 948 → the following by Petroff'sequation: (1) Friction coefficient for the bearing (2) Heat energy lost
 949 → due to friction in the bearing.
 950 Options: ['Friction coefficient for the bearing is 0.003284, Heat energy lost due to friction in the bearing
 951 → is 0.066 hp', 'Friction coefficient for the bearing is 0.005124, Heat energy lost due to friction in
 952 → the bearing is 0.070 hp', 'Friction coefficient for the bearing is 0.005784, Heat energy lost due
 953 → to friction in the bearing is 0.065 hp', 'Friction coefficient for the bearing is 0.005234, Heat
 954 → energy lost due to friction in the bearing is 0.072 hp', 'Friction coefficient for the bearing is
 955 → 0.002754, Heat energy lost due to friction in the bearing is 0.060 hp', 'Friction coefficient for
 956 → the bearing is 0.006234, Heat energy lost due to friction in the bearing is 0.075 hp', 'Friction
 957 → coefficient for the bearing is 0.003964, Heat energy lost due to friction in the bearing is 0.073 hp
 958 → ', 'Friction coefficient for the bearing is 0.007314, Heat energy lost due to friction in the
 959 → bearing is 0.082 hp', 'Friction coefficient for the bearing is 0.004564, Heat energy lost due to
 960 → friction in the bearing is 0.071 hp', 'Friction coefficient for the bearing is 0.004264, Heat energy
 961 → lost due to friction in the bearing is 0.068 hp'][br/>
 962 (Question 12) question: A football team entered into a 10-year lease with a city for use of the city's
 963 → athletic stadium. Five years into the lease, the team threatened to leave the stadium and move to
 964 → another city. The city sued the team in federal court, seeking a permanent injunction to prevent the
 965 → team from breaching its lease and leaving. In its answer, the team included a counterclaim seeking
 966 → \$10 million in damages for losses caused by the city's alleged failure to properly maintain the
 967 → stadium, as the lease required. The team demanded a jury trial on the counterclaim. The city moved to
 968 → try its claim for a permanent injunction before the trial on the team's counterclaim. The team
 969 → objected and moved that the jury trial of its counterclaim be held before the trial of the city's
 970 → injunction claim. How should the court rule on the parties' motions?
 971 Options: ['The court should first hold a nonjury trial of the city's claim without giving binding effect to
 972 → its findings or conclusions in the later jury trial of the team's counterclaim.', 'The court should
 973 → first hold a jury trial of the team's counterclaim, and its findings should be binding in the later
 974 → nonjury trial of the city's claim.', 'The court should first hold a nonjury trial of the team's
 975 → counterclaim, and then a jury trial of the city's claim.', 'The court should schedule a jury trial
 976 → of both the city's claim and the team's counterclaim.', 'The court should hold a simultaneous jury
 977 → trial of the team's counterclaim and nonjury trial of the city's claim.', 'The court should first
 978 → hold a jury trial of the city's claim, and then a nonjury trial of the team's counterclaim.', 'The
 979 → court should first hold a jury trial of the team's counterclaim, and then a nonjury trial of the
 980 → issues remaining in the city's claim.', 'The court should schedule a nonjury trial of both the city's
 981 → claim and the team's counterclaim.', 'The court should first hold a nonjury trial of the city's
 982 → claim, and then a jury trial of the issues remaining in the team's counterclaim.', 'The court should
 983 → first hold a nonjury trial of the city's claim, and its findings should be binding in the later
 984 → jury trial of the team's counterclaim.'][br/>
 985 (Question 13) question: Under which one of the following situations would the defendant(s) most likely be
 986 → found guilty of larceny?
 987 Options: ['A defendant, a trusted employee, was given the keys to the safe at the business where he worked.
 988 → One night, he took money from the safe, intending to return it after he won at the horse races.', 'A
 989 → defendant found an envelope full of money on a park bench. Knowing it was probably lost property,
 990 → he decided to keep it instead of turning it in to the police.', 'A defendant who was a contractor
 991 → took payment from a homeowner for materials and then used those materials on a different job.', 'A
 992 → During a craps game in the back room of a bar, the defendant lost \$150 to the victim. The defendant
 993 → left the bar after losing his money and returned an hour later with a gun in his possession.
 994 → Honestly believing that the \$150 still belonged to him, the defendant confronted the victim in the
 995 → back room and demanded the return of his money. Frightened, the victim handed the money back to the
 996 → defendant.', 'As two defendants were walking down the street, they noticed a victim park his
 997 → metallic gold sports car and enter the pool hall. When they approached the car, one of the
 998 → defendants observed that the victim had left the keys in the ignition. Knowing that the victim would
 999 → be hustling pool all evening, the defendants hopped into the sports car and drove off, intending to
 1000 → return the car later that evening.', 'One afternoon, the defendant noticed the victim riding his
 1001 → racing bike in the park. The defendant, who always wanted to own a racing bike, saw his opportunity
 1002 → when the victim left his bike unattended to participate in a touch football game. The defendant
 1003 → jumped on the bike and quickly rode away. Later that evening, the defendant called the victim and
 1004 → demanded \$200 for the return of the bike. The victim agreed to the defendant's demand. The following
 1005 → day, the victim paid the defendant the money, and the defendant, in turn, returned the bike.'][br/>
 1006 (Question 14) question: For a given level of government spending the federal government usually experiences a
 1007 → budget ____ during economic ____ and a budget ____ during economic ____.
 1008 Options: ['surplus expansion surplus expansion', 'deficit recession deficit expansion', 'surplus expansion
 1009 → deficit recession', 'deficit expansion deficit recession', 'surplus expansion surplus recession',
 1010 → deficit expansion surplus recession', 'surplus recession deficit recession'][br/>

972 → deficit recession surplus expansion', 'surplus recession deficit expansion', 'surplus recession
 973 → surplus expansion', 'deficit expansion surplus recession', 'surplus recession deficit recession']
 974 (Question 15) question: Let an undirected graph G with edges $E = \{<0,2>, <2,4>, <3,4>, <1,4>\}$, which $\langle A, B \rangle$
 975 → represent Node A is connected to Node B. What is the minimum vertex cover of G if 0 is one of vertex
 976 → cover? Represent the vertex cover in a list of ascending order.
 977 Options: ['[0, 1, 2]', '[0, 2, 4]', '[0, 1, 4]', '[0, 4]', '[0, 3, 4]', '[0, 1, 3]', '[0, 1]', '[0, 2, 3]',
 978 → '[0, 2]', '[0, 3]']
 979 (Question 16) question: A 35-year-old man with some experience as a truck driver owned a lumber truck. One day
 980 → , the man set out driving his truck, heavily loaded with lumber, down a mountain road. Sitting next
 981 → to the man in the passenger seat was a 19-year-old helper. During the course of the trip, when the
 982 → truck was going down long hill, the brakes failed. The man shouted to the helper to jump, but the
 983 → teenager refused to do so and shouted back to the man that he should try to steer the truck down the
 984 → hill. The man then opened the door on the passenger's side of the truck and negligently pushed the
 985 → helper out. The helper, who suffered a broken leg, was rushed to the hospital where he was treated
 986 → for his injury. As the helper was recuperating, a nurse inadvertently mixed up his chart with that of
 987 → the female patient in the next room. The nurse gave the helper a fertility pill that made him
 988 → sterile. If the helper brings suit against the man to recover damages for his sterility, the man will
 989 → Options: ['not recover, because the man had no control over the hospital's actions.', 'recover, because the
 990 → man failed to properly instruct the helper on emergency procedures.', 'not recover, because the man'
 991 → 's actions were intended to save the helper's life.', 'not recover, because the nurse's negligence in
 992 → giving the wrong medication is unrelated to the man's actions.', 'recover, because the man was
 993 → responsible for the helper's safety while in the truck.', 'recover, because the man was negligent in
 994 → his maintenance of the truck's brakes.', 'recover, because the man was negligent in pushing the
 995 → helper out of the truck.', 'not recover, because it is not foreseeable that a patient would be
 996 → injured in such a manner.', 'recover, because it is foreseeable that a hospital can be negligent in
 997 → its care of patients.'][Question 17) question: A note bearing interest at 7% for 60 days with a face value of \$3,000 was issued on
 998 → September 25. On November 1, the note was discounted at 5 (1/2)%. What were the proceeds?
 999 Options: ['\$3035', '\$2,985.75', '\$3,000', '\$3024.34', '\$210', '\$2,950', '\$3,015.50', '\$10.66', '\$3,100', '\$3
 1000 → ,050']
 1001 (Question 18) question: According to Socrates, the value or quality of one's life depends on understanding the
 1002 → principles of, or basic rationale for human existence. Without such knowledge (he suggests) life
 1003 → lacks virtue, because:
 1004 Options: ['understanding existence philosophically leads to personal satisfaction.', 'one's life lacks virtue
 1005 → if they do not contribute to society.', 'ignorance is a virtue in itself, as it leads to humility
 1006 → .', 'philosophical understanding is not necessary as long as one is happy.', 'acting virtuously
 1007 → means acting in way that is informed about what one is doing and why.', 'virtue is irrelevant to the
 1008 → quality of life.', 'someone who does not understand existence philosophically could never do
 1009 → anything right.', 'not only is virtue knowledge but also the unexamined life is not worth living.',
 1010 → "the value of life is determined by one's personal wealth and status.", 'to have the power or
 1011 → ability to do anything at all requires that we know what we are doing.'][Question 19) question: Octane burns in theoretical air (21% O₂, 79% N₂) by the following combustion equation
 1012 → : C₈H₁₈ + 12.5O₂ + 12.5(3.76)N₂ = 8CO₂ + 9H₂O + 47N₂ What is the theoretical air-fuel ratio?
 1013 Options: ['14.7g air/g fuel', '15.1g air/g fuel', '13.4g air/g fuel', '20.3g air/g fuel', '28.96g air/g fuel
 1014 → ', '10.0g air/g fuel', '18.5g air/g fuel', '12.5g air/g fuel', '16.2g air/g fuel', '47g air/g fuel']
 1015 (Question 20) question: The Shang Dynasty laid the foundation for a coherent development of Chinese
 1016 → civilization that lasted well into the:
 1017 Options: ['17th century.', '14th century.', '18th century.', '21st century.', '5th century.', '13th century
 1018 → .', '20th century.', '8th century.', '15th century.', '10th century.'][Question 21) question: Find the contributions of the conductivities of individual ions (ionic equivalent
 1019 → conductance λ_{+} and λ_{-}) when (equivalent conductance of electrolytic solution) =
 1020 → 0.0412 m² / (Ω) (mol) and $t_{+} = 0.825$.
 1021 Options: [' $\lambda_{+} = 0.00721 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.0412 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.825$
 1022 → m² / () (mol)', ' $\lambda_{-} = 0.0412 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.0340 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.0206 \text{ m}^2$
 1023 → / () (mol)', ' $\lambda_{+} = 0.0072 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.0307 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.0105 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.03295 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.0411 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.0001 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.0412 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.03399 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.03399 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.00721 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{+} = 0.0410 \text{ m}^2 / () \text{ (mol)}$ ', ' $\lambda_{-} = 0.0002 \text{ m}^2 / () \text{ (mol)}$ '][Question 22) question: Calculate the density of a block of wood which weighs 750 kg and has the dimensions 25
 1024 → cm 0.10 m 50.0 m.
 1025 Options: ['1.00 g/cc', '.70 g/cc', '.60 g/cc', '0.40 g/cc', '0.45 g/cc', '.80 g/cc', '0.30 g/cc', '.50 g/cc',
 1026 → '0.75 g/cc', '0.90 g/cc']
 1027 (Question 23) question: Use the van der Waals parameters for chlorine to calculate approximate values of the
 1028 → Boyle temperature of chlorine.
 1029 Options: ['\$1.25 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.20 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.41 \times 10^3 \text{ } \\$\text{mathrm{K}}\$',
 1030 → '\$1.05 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.10 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.65 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.75 \times
 1031 → 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.33 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.60 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.60 \times 10^3 \text{ } \\$\text{mathrm{K}}\$',
 1032 → '\$1.05 \times 10^3 \text{ } \\$\text{mathrm{K}}\$', '\$1.50 \times 10^3 \text{ } \\$\text{mathrm{K}}\$'][Question 24) question: In which instance would a state, under the enabling clause of the Fourteenth Amendment
 1033 → , be most able to regulate?
 1034 Options: ['A federal official from discriminating against a person based on race.', 'A federal official from
 1035 → discriminating against a person based on gender.', 'A federal official from discriminating against a
 1036 → person based on nationality.', 'A private company from discriminating against a person based on
 1037 → nationality.', 'A private individual from discriminating against a person based on race.', 'A state
 1038 → official from discriminating against a person based on race.', 'A private individual from discriminating against
 1039 → a person based on nationality.', 'A private company from discriminating against a person based on
 1040 → race.', 'A state official from discriminating against a person based on nationality.'][Question 25) question: Suppose a commercial bank has deposits of \$400,000 and has made loans and investments
 1041 → of \$315,000. Assume also that deposits are its only source of reserves. If the required reserve ratio
 1042 → is 20%, how much are its excess reserves?
 1043 Options: ['\$85,000', '\$80,000', '\$5,000', '\$315,000']
 1044 (Question 26) question: What percent of the total volume of an iceberg floats above the water surface? Assume
 1045 → the density of ice to be 57.2lb_m/ft³, the density of water 62.4lb_m/ft³.
 1046 Options: ['5 percent', '40 percent', '35 percent', '10 percent', '20 percent', '8 percent', '25 percent', '30
 1047 → percent', '12 percent', '15 percent']
 1048 (Question 27) question: If the mean rate of oxygen consumption of a male athlete during a training session is
 1049 → 2 l/min, then his rate of energy expenditure is approximately:
 1050 Options: ['20 kJ/min.', '500 kJ/min.', '30 kJ/min.', '400 kJ/min.', '60 kJ/min.', '200 kJ/min
 1051 → .', '100 kJ/min.', '80 kJ/min.', '120 kJ/min.'][Page-Footer] 19

1026 (Question 28) question: For most children, stranger anxiety begins at about ____ months of age.
 1027 Options: ['4 to 6', '11 to 13', '8 to 10', '2 to 3', '3 to 5', '15 to 18', '1 to 2', '12 to 14', '7 to 9', '6
 1028 → to 7']
 1029 (Question 29) question: In which instance would a state, under the enabling clause of the Fourteenth Amendment
 1029 → be most able to regulate?
 1030 Options: ['A federal official from discriminating against a person based on race.', 'A federal official from
 1031 → discriminating against a person based on gender.', 'A federal official from discriminating against a person based on
 1031 → nationality.', 'A private company from discriminating against a person based on
 1032 → nationality.', 'A private individual from discriminating against a person based on race.', 'A state
 1032 → official from discriminating against a person based on race.', 'A private individual from
 1033 → discriminating against a person based on gender.', 'A private individual from discriminating against
 1033 → a person based on nationality.', 'A private company from discriminating against a person based on
 1034 → race.', 'A state official from discriminating against a person based on nationality.'][br/>
 1034 (Question 30) question: A 360 → journal bearing 3 in.long, carries a 4 in. diameter shaftwith a radial
 1035 → clearance of 0.0025 in. The shaft supports aradial load of 1000 lbs. and at a speed of 500 rpm. The
 1036 → operatingtemperature for the bearing is 140 Fand SAE 20 oilis used for bearing lubrication. Evaluate
 1036 → the following by Petroff'sequation: (1) Friction coefficient for the bearing (2) Heat energy lost
 1037 → due to friction in the bearing.
 1038 Options: ['Friction coefficient for the bearing is 0.003284, Heat energy lost due to friction in the bearing
 1038 → is 0.066 hp', 'Friction coefficient for the bearing is 0.005124, Heat energy lost due to friction in
 1039 → the bearing is 0.070 hp', 'Friction coefficient for the bearing is 0.005784, Heat energy lost due
 1039 → to friction in the bearing is 0.065 hp', 'Friction coefficient for the bearing is 0.005234, Heat
 1040 → energy lost due to friction in the bearing is 0.072 hp', 'Friction coefficient for the bearing is
 1040 → 0.002754, Heat energy lost due to friction in the bearing is 0.060 hp', 'Friction coefficient for
 1041 → the bearing is 0.006234, Heat energy lost due to friction in the bearing is 0.075 hp', 'Friction
 1041 → coefficient for the bearing is 0.003964, Heat energy lost due to friction in the bearing is 0.073 hp
 1042 → ', 'Friction coefficient for the bearing is 0.007314, Heat energy lost due to friction in the
 1042 → bearing is 0.082 hp', 'Friction coefficient for the bearing is 0.004564, Heat energy lost due to
 1043 → friction in the bearing is 0.071 hp', 'Friction coefficient for the bearing is 0.004264, Heat energy
 1044 → lost due to friction in the bearing is 0.068 hp']
 1045
 1046 What construct is present in the positive examples but absent in the negative examples? Provide a concise
 1046 → label and brief explanation.
 1047
 1048 **D.2 PROMPT FOR INTERPRETING CLUSTERS**
 1049
 1050
 1051 (2) Analyze the following positive and negative examples to identify the main theme that appear in positive
 1051 → examples and not in negative examples.
 1052
 1053 POSITIVE EXAMPLES (items that load on this dimension):
 1053 (Question 1) question: What is the major difference between naturalism and positivism?
 1054 Options: ['Naturalism concerns the underpinning values of the international society, while positivism the
 1054 → international rules as such', 'Naturalism is a contemporary theory, while positivism is a
 1055 → traditional theory', 'On the one hand, naturalism is based on a set of rules that are of universal
 1055 → and objective scope and on the other hand, positivism is based on a structured and coherent legal
 1056 → system that is created by States in light of their interests and desires', 'Naturalism argues that
 1056 → laws are inherently subjective and influenced by personal beliefs, while positivism argues that laws
 1057 → are objective and universally applicable', 'Naturalism focuses on the interpretation of law, while
 1057 → positivism focuses on the enforcement of law', 'Naturalism is a theory of law that is subjective and
 1058 → varies from person to person, while positivism is a theory of law that is objective and the same
 1058 → for everyone.', 'Naturalism is a school of thought advocated outside of Europe, while positivism is
 1059 → Euro-centric theory', 'Naturalism is a theory of law that relies on the inherent morality of a
 1060 → society, while positivism relies on the codification of laws by a governing body', 'While naturalism
 1061 → serves only the theory human rights law, positivism has a wider and more general scope', 'Naturalism
 1062 → is concerned with the moral aspects of law, while positivism is concerned with the legal
 1062 → aspects of law']
 1063 (Question 2) question: Which among the following prevents overfitting when we perform bagging?
 1063 Options: ['The use of all data without sampling', 'The practice of validation performed on a random subset of
 1064 → classifiers trained', 'The use of underfitting to counterbalance overfitting', 'The use of weak
 1064 → classifiers', 'The use of classification algorithms which are prone to overfitting', 'The use of
 1065 → classification algorithms which are not prone to overfitting', 'The use of sampling with replacement
 1065 → as the sampling technique', 'The practice of validation performed on every classifier trained', 'The
 1066 → use of strong classifiers', 'The use of sampling without replacement as the sampling technique']
 1067 (Question 3) question: A witness lived next door to a victim. Late one night, the witness overheard the victim
 1067 → scream, "The defendant, please don't do it!" The next day the victim was found dead in her bedroom.
 1068 → The defendant was arrested and charged with murdering the victim. At trial, the witness proposes to
 1068 → testify to the victim's statement. Upon objection by defendant's counsel, the court should rule the
 1069 → witness's testimony regarding the victim's statement
 1070 Options: ['admissible as a past recollection recorded if the witness can confirm the statement was made.',
 1070 → 'admissible as a dying declaration if the jury determines that the victim believed that her death was
 1071 → imminent.', 'admissible as a dying declaration if the judge, by preponderance of the evidence,
 1071 → determines that the victim believed that her death was imminent.', 'inadmissible, because the
 1072 → probative value is substantially outweighed by the danger of unfair prejudice.', 'inadmissible,
 1072 → because the statement was not made under oath.', 'inadmissible as hearsay not within any recognized
 1073 → exception.', 'inadmissible, because the statement is considered a character evidence.', 'admissible,
 1073 → because the statement is a present sense impression.', 'inadmissible, because the statement can
 1074 → lead to a character attack on the defendant.'][br/>
 1075 (Question 4) question: Suppose that $f(x)$ is a polynomial that has degree 6 and $g(x)$ is a polynomial that
 1075 → has degree 3. If $h(x)$ is also a polynomial such that $f(g(x)) + g(h(x)) + h(f(x))$ is a
 1076 → polynomial of degree 36, then what is the degree of the polynomial h ?
 1077 Options: ['3', '21', '6', '9', '30', '72', '12', '15', '24', '18']
 1077 (Question 5) question: What happens to the image formed by a concave mirror as the object is moved from far
 1078 → away to near the focal point?
 1078 Options: ['The image disappears as it moves toward the mirror.', 'The image moves away from the mirror and
 1078 → gets taller.', 'The image stays at the same distance from the mirror and gets shorter.', 'The image
 1079 → moves toward the mirror and gets shorter.', 'The image moves toward the mirror and gets taller.', 'The
 1079 → image stays at the same distance from the mirror and gets taller.'][

1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110
 1111
 1112
 1113
 1114
 1115
 1116
 1117
 1118
 1119
 1120
 1121
 1122
 1123
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133

→ mirror and stays the same height.', 'The image flips upside down as it moves toward the mirror.', '
 → The image moves away from the mirror and stays the same height.', 'The image moves away from the
 → mirror and gets shorter.]

(Question 6) question: Which of the following creates the trade-off depicted by the Phillips curve?
 Options: ['A decrease in input costs that corresponds with a decrease in unemployment', 'A decrease in output
 → that corresponds with an increase in the price level', 'An increase in input costs that corresponds
 → with an increase in unemployment', 'A rightward shift in the aggregate demand curve', 'An increase
 → in output that corresponds with a decrease in the price level', 'A decrease in input costs that
 → corresponds with an increase in output', 'An increase in output that corresponds with a decrease in
 → unemployment', 'A leftward shift in the aggregate supply curve', 'A leftward shift in the aggregate
 → demand curve', 'A rightward shift in the aggregate supply curve']

(Question 7) question: A sleeve-coupling is used on a shaft 1(1/2) in. in diameter, delivering a torque of 500
 → lbs.-in. Calculate the diameter of the pin required to hold the coupling if the design stress for
 → the pin material in shear is 15,000 psi.
 Options: ['0.145 in.', '0.125 in.', '0.210 in.', '0.155 in.', '0.162 in.', '0.168 in.', '0.190 in.', '0.158
 → in.', '0.175 in.', '0.182 in.'])

(Question 8) question: The Supreme Court has original jurisdiction in certain areas. Congress has given lower
 → federal courts concurrent jurisdiction with the Supreme Court in all cases except for disputes
 → involving:
 Options: ['maritime issues', 'private citizens', 'international disputes', 'federal laws', 'different states
 → ', 'constitutional matters', 'a state as a party', 'federal officials', 'public ministers and
 → consuls']

(Question 9) question: Which of the following Internet-related challenges is most significant in the public
 → relations field?
 Options: ['Overcoming language barriers in global communications', 'Staying abreast of changing technology',
 → 'Ensuring security of online client information', 'Adapting to the rapid advancement of digital
 → marketing tools', 'Finding stable, cost-effective Internet providers', 'Representing clients using
 → new social media environments', 'Training staff to use social media']

(Question 10) question: Archaeological evidence for the collapse of civilizations suggests which of the
 → following is the most important variable?
 Options: ['the existence of a written language', 'the number of years the civilization has existed', "the
 → civilization's religious beliefs and practices", 'whether warfare can be ended', "the size of the
 → civilization's population", 'the ability of bureaucracies to store food and feed its people', "the
 → civilization's trade networks with other societies", 'changing environmental conditions', 'the
 → development of new technologies', 'how societies respond to challenges']

NEGATIVE EXAMPLES (items that do NOT load on this dimension):

(Question 1) question: Given that Bismuth has a density of $9.80 \times 10^3 \text{ kg m}^{-3}$ and a mass absorption
 → coefficient for a gamma radiation corresponding to a half thickness of 12 g cm^{-2} . What is the percent
 → decrease in intensity at a depth of 1.5 cm?
 Options: ['75%', '50%', '55%', '80%', '65%', '45%', '57.15%', '40%', '70%', '60%']

(Question 2) question: How can one demonstrate that amino acids are used to build proteins in cells, while the
 → proteins already in the cell are used in catabolism?
 Options: ['Enzyme-linked immunosorbent assay (ELISA)', 'Polymerase chain reaction (PCR)', 'X-ray
 → crystallography', 'Mass spectrometry', 'Autoradiography', 'Western blotting', 'Microscopy', 'Chromatography', 'Fluorescence in situ hybridization (FISH)', 'Spectroscopy']

(Question 3) question: What are the rules regarding the specification of a procedure?
 Options: ['Procedures execute automatically during program execution', 'Procedures require a start statement
 → but do not have a specific ending requirement; they terminate when the program execution completes
 → .', 'Procedures can only be declared at the beginning of the program', 'Procedure blocks must be
 → written in a separate file from the main program and are included during compilation.', 'Procedures
 → can be declared with any number of names, and the END statement is optional, with the block ending
 → at the first RETURN statement.', 'Procedures must be invoked at the start of the program and run
 → continuously until the program ends.', 'Procedure blocks must be labeled with a unique number and
 → can only contain variable declarations.', 'Procedures do not require an end statement', 'Procedures
 → must be declared within the main function and cannot be invoked by other procedures or functions.',
 → 'Procedure block may appear anywhere in the program, must start with a procedure statement with at
 → least one name, must end with an END procedure-name statement, is bypassed during execution unless
 → invoked by a procedure reference.]

(Question 4) question: A defendant wished to see his high school basketball team win the state championship.
 → During an important game, the defendant pulled out a gun and shot at the leg of a key player on the
 → opposing team. The defendant intended only to inflict a slight wound so that the opposing player
 → would be unable to complete the game. When the defendant fired the shot, he unintentionally hit a
 → player on his own high school team in the chest, killing him instantly. What is the most serious
 → crime that the defendant can be convicted of?
 Options: ['Assault with a deadly weapon.', 'Battery.', 'Involuntary manslaughter.', 'Voluntary manslaughter
 → .', 'Murder.', 'Felonious assault.', 'Reckless endangerment.', 'Justifiable homicide.', 'Attempted
 → murder.', 'Negligent homicide.'])

(Question 5) question: A landowner, her aunt, and her brother are the owners of three contiguous lots. A
 → downward slope exists from the landowner's land to the brother's land. The aunt, the owner in the
 → middle of the three lots, was an amateur scientist. She began experimentation with electricity, using
 → a lightning rod to conduct her experiments. Frequently, the electrical storms apparently originated
 → in and over the area within which the brother's land is located. Because of this, the aunt would
 → often angle her rod over the fence between her property and the brother's property to maximize her
 → chances of success. After one entire rainy and windy season of experiments, the brother had grown
 → tired of the aunt's lightning-rod intrusions because they interfered with his ham radio operations in
 → his garage. The brother brought suit to enjoin such lightning-rod experiments. At trial, it was
 → found that the aunt had been conducting her activities in the airspace directly above the brother's
 → land. The court should
 Options: ['not enjoin the aunt's experiments because the brother does not own the space above his land.', "enjoin the aunt's experiments because they can potentially cause harm to the brother's ham radio
 → operations.", "enjoin the aunt's experiments because they constitute an unreasonable interference
 → with the space above the brother's land.", "enjoin the aunt's experiments because one does not have
 → the right to engage in such scientific experiment without permission.", "enjoin the aunt's
 → experiments because the brother has a right to quiet enjoyment of his property.", "not enjoin the
 → aunt's experiments if they do not physically damage the brother's property.", "not enjoin the aunt's
 → experiments because they contribute to scientific research and discovery.", "enjoin the aunt's
 → experiments because she did not have proper permits for scientific experimentation.", "not enjoin
 → the aunt's lightning rod experiments if they were necessary to protect her property from the
 → electrical storms.", "not enjoin the aunt's experiments because the electrical storms are natural
 → occurrences that she cannot control."]'

1134 (Question 6) question: Maximize the entropy $H(X)$ of a non-negative integer-valued random variable $\$X\$$,
 1135 → taking values from 0 to infinity, subject to the constraint $\$E(X)=1\$$. Use base 2 logarithm to
 1136 → evaluate $\$H(X)\$$.
 1137 Options: ['4.0', '2.0', '3.0', '1.0', '4.5', '2.5', '3.5', '0.5', '5.0', '1.5']
 1138 (Question 7) question: Copper has a conductivity $\sigma = 5.8 \times 10^{-7} \text{ mhos/meter}$, and $\mu = 1.26 \times 10^{-6}$
 1139 → Henry/meter. If a plane wave at a frequency of 200 Hz is normally incident on copper find the depth
 1140 → of penetration of this wave.
 1141 Options: ['5.77 \$10^{-3} \text{ meter}', '6.22 \$10^{-3} \text{ meter}', '3.50 \$10^{-3} \text{ meter}', '2.80 \$10^{-3} \text{ meter}', '5.00
 1142 → \$10^{-3} \text{ meter}', '7.34 \$10^{-3} \text{ meter}', '8.20 \$10^{-3} \text{ meter}', '9.50 \$10^{-4} \text{ meter}', '4.66 \$10^{-3}
 1143 → meter', '1.58 \$10^{-2} \text{ meter}']
 1144 (Question 8) question: What is the smallest refracting angle that a glass prism ($n = 1.5$) can have so that no
 1145 → ray can be transmitted through it? What is this angle for a water prism ($n = 1.33$)?
 1146 Options: ['85 31° 16" for a glass prism and 99 17° 51" for a water prism', '81 24° 12" for a glass
 1147 → prism and 95 13° 43" for a water prism', '87 29° 15" for a glass prism and 101 19° 53" for
 1148 → a water prism', '83 27° 13" for a glass prism and 97 14° 46" for a water prism', '86 28°
 1149 → 14" for a glass prism and 100 18° 52" for a water prism', '82 25° 12" for a glass prism and 96
 1150 → 13° 42" for a water prism', '82 23° 10" for a glass prism and 96 11° 41" for a water prism
 1151 → ', '80 22° 11" for a glass prism and 94 12° 40" for a water prism', '84 30° 15" for a glass
 1152 → prism and 98 16° 50" for a water prism', '81 26° 14" for a glass prism and 95 15° 48" for
 1153 → a water prism']
 1154 (Question 9) question: If $v(t) = 10e^{-4t} u(t)$, and $v_{ab}(s)$ is related to $V(s)$ by $v_{ab}(s) = \{3V(s) + 3s + 27\} / (s^2 + 6s + 8)$, formulate $v_{ab}(t)$.
 1155 Options: [' $v_{ab}(t) = -15e^{-4t} - 15e^{-4t} + 18e^{-2t}$ ', ' $v_{ab}(t) = 15e^{-4t} - 15e^{-4t} + 18e^{-2t}$ ', ' $v_{ab}(t) = 30$
 1156 → $te^{-4t} + 30e^{-4t} - 36e^{-2t}$ ', ' $v_{ab}(t) = 30te^{-4t} - 30e^{-4t} + 36e^{-2t}$ ', ' $v_{ab}(t) = 15te^{-4t} + 15e^{-4t} - 18e^{-2t}$ ',
 1157 → ' $v_{ab}(t) = 15te^{-4t} + 15e^{-4t} + 18e^{-2t}$ ', ' $v_{ab}(t) = -15te^{-4t} - 15e^{-4t} - 18e^{-2t}$ ',
 1158 → ' $v_{ab}(t) = -15te^{-4t} + 15e^{-4t} + 18e^{-2t}$ ', ' $v_{ab}(t) = -30te^{-4t} - 30e^{-4t} + 36e^{-2t}$ ']
 1159 (Question 10) question: If all the values of a data set are the same, all of the following must equal zero
 1160 → except for which one?
 1161 Options: ['Mean', 'Standard deviation', 'Variance', 'Interquartile range', 'Quartile deviation', 'Range',
 1162 → 'Coefficient of variation', 'Percentile']
 1163 (Question 11) question: A nation is producing at a point inside of its production possibility curve. Which of
 1164 → the following is a possible explanation for this outcome?
 1165 Options: ['This nation has experienced a permanent decrease in its production capacity.', 'This nation has
 1166 → experienced slower than usual technological progress.', 'This nation has avoided free trade between
 1167 → other nations.', 'This nation is experiencing an economic recession.'][br/>
 1168 (Question 12) question: Let $\$p\$$, $\$q\$$, and $\$r\$$ be constants. One solution to the equation $\$(x-p)(x-q) = (r-p)(r-q)\$$
 1169 → is $\$x=r\$$. Find the other solution in terms of $\$p\$$, $\$q\$$, and $\$r\$$.
 1170 Options: [' $r-p-q'$, ' $p+q-r'$ ', ' $pq-r'$ ', ' pqr' ', ' $p*q/r'$ ', ' $p*r/p'$ ', ' $(p*q)/r'$]
 1171 (Question 13) question: The atomic radius of hydrogen is 0.037 nm. Compare this figure with the length of the
 1172 → first Bohr radius. Explain any differences.
 1173 Options: ['The atomic radius of hydrogen is five times the first Bohr radius', 'The atomic radius of hydrogen
 1174 → is four times the first Bohr radius', 'The atomic radius of hydrogen is half the first Bohr radius
 1175 → ', 'The atomic radius of hydrogen is one tenth the first Bohr radius', 'The atomic radius of
 1176 → hydrogen is larger than the first Bohr radius', 'The atomic radius of hydrogen is unrelated to the
 1177 → first Bohr radius', 'The atomic radius of hydrogen is smaller than the first Bohr radius', 'The
 1178 → atomic radius of hydrogen is three times the first Bohr radius', 'The atomic radius of hydrogen is
 1179 → twice the first Bohr radius', 'The atomic radius of hydrogen is equal to the first Bohr radius']
 1180 (Question 14) question: If a binary component system containing initially 3g of component A and 7g of
 1181 → component B is brought to a temper-ature such that two phases are present, one with a wt % of
 1182 → component A = 5% and the other with wt % of component A = 50%, calculate the masses of the two phases
 1183 → .
 1184 Options: [' $m_{\alpha} = 4.444g$, $m_{\beta} = 6.556g$ ', ' $m_{\alpha} = 5.000g$, $m_{\beta} = 5.000g$ ', ' $m_{\alpha} = 6.000g$, $m_{\beta} =$
 1185 → $4.000g$ ', ' $m_{\alpha} = 2.500g$, $m_{\beta} = 7.500g$ ', ' $m_{\alpha} = 3.000g$, $m_{\beta} = 7.000g$ ', ' $m_{\alpha} =$
 1186 → $1.500g$, $m_{\beta} = 8.500g$ ', ' $m_{\alpha} = 5.556g$, $m_{\beta} = 4.444g$ ', ' $m_{\alpha} = 4.444g$, $m_{\beta} = 5.556g$ ',
 1187 → ' $m_{\alpha} = 3.444g$, $m_{\beta} = 6.556g$ ', ' $m_{\alpha} = 2.222g$, $m_{\beta} = 7.778g$ ']
 1188 (Question 15) question: Eighty-five more than the square of a number is the same as the square of the quantity
 1189 → that is \$17\$ less than the number. What is the number?
 1190 Options: ['17', '11', '12', '5', '8', '2', '3', '10', '7', '6']
 1191 (Question 16) question: Which one of the following is NOT an advantage of urban agriculture?
 1192 Options: ['Beautification of a dingy urban area', 'Improved air quality', 'Increase in property values',
 1193 → 'Helping to solve the problem of solid waste disposal', 'Increased urban biodiversity', 'Reduction in
 1194 → carbon footprint', 'Renewed or purified water supplies', 'Fresh produce for sale to others',
 1195 → 'Reduced dependence on fossil fuels']
 1196 (Question 17) question: China and Vietnam's dispute over the Spratley Islands is
 1197 Options: ['a religious dispute.', 'a resource dispute.', 'a cultural dispute.', 'a historical dispute.', 'a
 1198 → functional dispute.'][br/>
 1199 (Question 18) question: Which of the following describes the cluster of blood capillaries found in each
 1200 → nephron in the kidney?
 1201 Options: ['Afferent arteriole', 'Glomerulus', 'Renal medulla', 'Proximal convoluted tubule', 'Efferent
 1202 → arteriole', 'Loop of Henle', 'Distal convoluted tubule', 'Renal pelvis', 'Bowman's capsule', 'Renal
 1203 → capsule']
 1204 (Question 19) question: In the 1930s F.T. Bacon designed hydrogen fuel cells with extremely good performance.
 1205 → Using his concepts Pratt and Whitney has constructed hydrogen fuel cells that produce a potential of
 1206 → 0.9 V and generate a current of about $2.3 \times 10^{-3} \text{ A}$ for each square meter of anode area. If a Bacon
 1207 → battery is to deliver 100 W of power, how large must the anode be?
 1208 Options: [' $3.2 \times 10^{-1} \text{ m}^2$ ', ' $5.0 \times 10^{-1} \text{ m}^2$ ', ' $2.2 \times 10^{-1} \text{ m}^2$ ', ' $7.8 \times 10^{-1} \text{ m}^2$ ', ' $1.0 \times 10^{-1} \text{ m}^2$ ', ' $2.8 \times 10^{-1} \text{ m}^2$
 1209 → ', ' $4.3 \times 10^{-1} \text{ m}^2$ ', ' $6.5 \times 10^{-1} \text{ m}^2$ ', ' $3.6 \times 10^{-1} \text{ m}^2$ ', ' $1.5 \times 10^{-1} \text{ m}^2$ ']
 1210 (Question 20) question: Consider the initial value problem
 1211
$$y'' + 2ay' + \left(a^2 + \frac{1}{4}\right)y = 0, \quad y(0) = 1, \quad y'(0) = 0.$$

 1212 For $\$a=1\$$ find the smallest $\$T\$$ such that $|y(t)| < 0.1\$$ for $\$t>T\$$.
 1213 Options: ['3.1415', '3.0000', '1.8763', '1.5708', '2.7183', '2.1345', '0.7854', '2.0000', '2.3754', '1.6520']
 1214 (Question 21) question: What is the measure of an angle that turns through $3/4$ of a complete circle?
 1215 Options: ['225°', '180°', '270°', '120°', '75°', '90°', '34°', '360°', '45°', '43°']
 1216 (Question 22) question: In theorising the Responsibility to Protect, fundamental questions are raised about
 1217 → the nature of state sovereignty. Regarding this tension between R2P and notions of sovereignty, which
 1218 → of the following is not an accurate statement?
 1219 Options: ['In order to transcend the intervention-vs.-sovereignty debate, R2P seeks a new definition of
 1220 → sovereignty.', 'In order to facilitate the deployment of effective humanitarian intervention, the
 1221 → R2P approach redefines the traditional Westphalian conception of sovereignty.', 'The notion
 1222 → conceptualised by the R2P approach is of a sovereignty that is intrinsically linked with

1188
 ↪ responsibilities and the duty to protect the interests of citizens.', 'According to R2P, sovereignty
 ↪ is only valid if the state can guarantee the safety and rights of its citizens.', 'In order to
 ↪ legitimise the inter-state interventions required by R2P, Westphalian notions of sovereignty are
 ↪ invoked, specifically the idea that intervention is required to assure international security.]'
 1191 (Question 23) question: Which of the following should be named using Roman numerals after the cation?
 Options: ['AlBr³', 'Na²⁰', 'K^{2Cr207}', 'CaCl²', 'NH₄Cl', 'H₂SO₄', 'MgSO₄', 'CuCl²', 'Zn(NO₃)²']
 1192 (Question 24) question: An owner operated a successful retail business. He decided he wanted to retire and
 ↪ listed his business for sale. A buyer saw the listing and visited the retail business. The buyer told
 ↪ the owner that she was very interested in buying the business, but that she needed to get a loan
 ↪ from the bank to finance the purchase. The parties drafted a contract, signed by both parties, which
 ↪ stated that the buyer agreed to purchase the retail business from the owner for \$250,000. The written
 ↪ agreement contained a provision wherein the sale would not take effect "unless the buyer is
 ↪ successful in her bid to obtain a loan from a bank to finance the purchase of the business."
 ↪ Subsequently, the buyer made no effort to attempt to obtain a loan from a bank so that the sale could
 ↪ be finalized. After several months, the owner filed suit against the buyer for breach of contract.
 ↪ Which of the following will be the owner's strongest argument in support of his action against the
 ↪ buyer for breach of contract?
 Options: ['The buyer's lack of effort in obtaining a loan is a breach of good faith and fair dealing.', 'The
 ↪ buyer was under a duty to inform the owner of her decision not to seek a loan.', 'Although obtaining
 ↪ a loan from a bank was a condition to the buyer's duty to buy the business, the condition should be
 ↪ excused because its non-occurrence would cause a forfeiture to the owner.', 'The owner suffered
 ↪ financial loss due to the buyer's inaction in seeking a loan.', 'The buyer breached an implied
 ↪ promise to make a reasonable effort to obtain a loan from a bank.', 'The obtaining of a loan from a
 ↪ bank was not worded as a "condition" to the buyer's duty to buy the business.', 'Although obtaining
 ↪ a loan from a bank was a condition to the buyer's duty to buy the business, it should be stricken
 ↪ from the contract because it is an unconscionable term.', 'The buyer did not explicitly state in the
 ↪ contract that she would not seek a loan.']
 1199 (Question 25) question: If your nominal income rises 4 percent and your real income falls 1 percent by how
 ↪ much did the price level change?
 Options: ['4 percent increase', '3 percent increase', 'No change in the price level', '5 percent decrease',
 ↪ '5 percent increase', '2 percent increase', '1 percent decrease', 'percent increase', '7 percent
 ↪ increase', '6 percent decrease']
 1207 (Question 26) question: What was the significance of the Truman Doctrine?
 Options: ['It outlined Truman's plan for economic recovery after World War II', 'It led to the establishment
 ↪ of the United Nations', 'It indicated that the US would now view all local revolts through a
 ↪ geopolitical lens', 'It was a declaration of war against the Soviet Union', 'It indicated the
 ↪ special place of Greece and Turkey in American interests', 'It signaled the US's withdrawal from
 ↪ international affairs', 'It marked the beginning of the Space Race', 'It marked the end of the Cold
 ↪ War', 'It was Truman's first statement on European affairs', 'It indicated US reluctance to get
 ↪ involved in incidents outside of its immediate sphere of influence']
 1212 (Question 27) question: A mixture of water and ethanol is prepared with a mole fraction of water of 0.60 . If
 ↪ a small change in the mixture composition results in an increase in the chemical potential of water
 ↪ by \$0.25 $\backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, by how much will the chemical potential of ethanol change?
 Options: ['0.15 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.15 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '0.60 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}', '0.25 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.25 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '0.38 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.20 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.75 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.38 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$, '-0.60 $ \backslash \text{mathrm}^{(J)} \backslash \text{mathrm}^{(\text{mol})^{(-1)}}$']
 1217 (Question 28) question: Lithium oxide (Li₂O, molecular weight = 30 g/mole) re-acts with water (H₂O,
 ↪ molecular weight = 18 g/mole, density = 1.0 g/cm³) to produce lithium hydroxide (LiOH) according to
 ↪ the following reaction: Li₂O + H₂O → 2LiOH. What mass of Li₂O is required to completely react
 ↪ with 24 liters of H₂O?
 Options: ['55 kg', '60 kg', '50 kg', '45 kg', '30 kg', '25 kg', '35 kg', '70 kg', '20 kg']
 1220 (Question 29) question: Which muscle is the most active during a right lateral excursion of the mandible?
 Options: ['Left masseter muscle', 'Right buccinator muscle', 'Right temporalis muscle', 'Right masseter
 ↪ muscle', 'Right lateral pterygoid muscle', 'Left medial pterygoid muscle', 'Left temporalis muscle',
 ↪ 'Left buccinator muscle', 'Left lateral pterygoid muscle', 'Right medial pterygoid muscle']
 1222 (Question 30) question: How many moles of Al₂O₃ can be formed when a mixture of 0.36 moles of aluminum and
 ↪ 0.36 moles of oxygen is ignited? Which substance and how much of it is in excess of that required? 4
 ↪ Al + 3O₂ → 2Al₂O₃
 Options: ['0.36 moles of Al₂O₃, 0.09 moles of O₂ in excess', '0.18 moles of Al₂O₃, 0.09 moles of O₂
 ↪ in excess', '0.18 moles of Al₂O₃, 0.18 moles of Al in excess', '0.30 moles of Al₂O₃, 0.06
 ↪ moles of O₂ in excess', '0.18 moles of Al₂O₃, no excess', '0.12 moles of Al₂O₃, 0.24 moles of
 ↪ O₂ in excess', '0.36 moles of Al₂O₃, no excess', '0.27 moles of Al₂O₃, no excess', '0.24
 ↪ moles of Al₂O₃, 0.12 moles of Al in excess', '0.27 moles of Al₂O₃, 0.09 moles of O₂ in excess
 ↪ ']
 1227 (Question 31) question: In the Tularosa Basin of New Mexico are black lava formations surrounded by light-
 ↪ colored sandy desert. Pocket mice inhabit both areas. Dark-colored ones inhabit the lava formations,
 ↪ while light-colored mice inhabit the desert. Which of the following statements is correct about this
 ↪ scenario?
 Options: ['The light-colored mice evolved from the dark-colored mice over time.', 'The color of the mice is
 ↪ determined by their diet in the lava formations and desert.', 'The two varieties of mice descended
 ↪ from a recent common ancestor.', 'The mouse population was originally one population that diverged
 ↪ into two species because of mutations.', 'Selection favors some phenotypes over others.', 'The two
 ↪ varieties of mice have always existed separately and did not share a common ancestor.', 'Originally
 ↪ the mice were all dark colored. As the lava decomposed into sand, the color of some mice changed
 ↪ because that color was favored in that environment.', 'The mice were originally light-colored, but
 ↪ some darkened over time to blend into the lava formations.'][br/>
 1234 (Question 32) question: The principal effect of the current concern to improve the employment opportunities of
 ↪ the educationally or socially disadvantaged has been to encourage the use in employment testing of
 Options: ['intelligence quotient tests', 'norm-referenced test', 'tests that assess physical abilities',
 ↪ work-sample types of tests', 'timed tests of a narrow range of abilities', 'personality assessment
 ↪ tests', 'tests based solely on academic performance', 'tests based on social skills']
 1237 (Question 33) question: Select the best translation into predicate logic: Either Taylor or Eric is a latte boy
 ↪ .
 Options: ['Lt Le', 'Lx Ly', 'Lt Le', 'L(t e)', 'tL eL', 'tL eL', 'Lx Ly', 'L(t
 ↪ e)', 'E T', 'tL eL']
 1239 (Question 34) question: How many trees are there on 5 unlabeled vertices?
 Options: ['3', '6', '8', '10', '11', '12', '15', '9', '7', '5']
 1240 (Question 35) question: The asteroid Pallas has an orbital period of 4.62 years and an orbital eccentricity of
 ↪ 0.233. Find the semi-major axis of its orbit. (Unit: 10¹¹ m)
 Options: ['5.55', '5.20', '6.25', '3.98', '4.76', '7.10', '3.14', '2.87', '4.15', '8.03']$

1242 (Question 36) question: In a double-slit experiment, $D = 0.1$ mm and $L = 1$ m. If yellow light is used, what
 1243 → will be the spacing between adjacent bright lines?
 1244 Options: ['3 mm', '9 mm', '1.5 mm', '12 mm', '4.5 mm', '15 mm', '7.5 mm', '10 mm', '6 mm', '18 mm']
 1245 (Question 37) question: Assume the half-life of the proton is 10^{33} years. How many decays per year would you
 → expect in a tank of water containing 350,000 liters of water?
 1246 Options: ['1.0', '0.015', '0.0008', '5.0', '0.5', '1.5', '0.08', '2.4', '3.0', '0.003']
 1247 (Question 38) question: Farmer Fred was the owner of Grapeview Farm, the largest farm in the County of
 → Grapeview. Fred was advised by his gift and estate attorney to begin making annual fractionalized
 → gifts of the property to his children. Fred executed a conveyance instrument, stating the gifts as "
 1248 → one-thirtieth of my ownership interest in Grapeview County, the eastern one tenth of the western
 → corner of the property is hereby given to all my children per capita." This description of the land
 → to be conveyed is
 1249 Options: ['Sufficient, because it gives a clear description of the fractional gift.', 'Sufficient, because it
 → mentions the name of the county and the property.', 'Sufficient, because it includes the name of
 → the property and a portion of it.', 'Insufficient, because it doesn't state the exact measurements
 → of the land to be given.', 'Insufficient, because the gifts to the children lacked consideration.',
 1250 → 'Insufficient, because it doesn't specify the exact location of the eastern one-tenth.',
 → 'Insufficient, because it is too vague.', 'Sufficient, because the property can be identified.',
 1251 → 'Sufficient, because it satisfies the title of seisin.'][br/>
 1252 (Question 39) question: Joe is the trustee of a trust set up for his father. Under the Internal Revenue Code,
 → when Joe prepares the annual trust tax return, Form 1041, he
 1253 Options: ['May sign the return but must include a separate statement explaining his relationship to the trust
 → .', 'May not sign the return unless he is also named as a co-trustee.', 'Can only sign the return if
 → he is also a certified public accountant.', 'Is considered a tax return preparer only if he is also
 → the beneficiary of the trust.', 'May not sign the return unless he receives additional compensation
 → for the tax return.', 'Must obtain the written permission of the beneficiary prior to signing as a
 → tax return preparer.', 'Is considered a tax return preparer because his father is the grantor of the
 → trust.', 'Is not considered a tax return preparer.', 'Is not considered a tax return preparer
 → unless he personally benefits from the trust's assets.', 'Is considered a tax return preparer and
 → must pay a fee to the IRS for this service.'][br/>
 1254 (Question 40) question: Knowing that the K_{sp} for AgCl is 1.8×10^{-10} , calculate E , the electrode potential
 1255 → for a silver-silver chloride electrode immersed in 1M KCl . The standard oxidation potential for
 → the (Ag, Ag^+) half reaction is -0.799 volts.
 1256 Options: ['0.576 volts', '0.344 volts', '-0.799 volts', '-0.344 volts', '0.799 volts', '-1.223 volts', '1.022
 → volts', '-0.576 volts', '-0.223 volts', '0.223 volts']
 1257 (Question 41) question: On either side of a pane of window glass, temperatures are 70 F and 0 F. How fast is
 → heat conducted through such a pane of area 2500 cm^2 if the thickness is 2 mm?
 1258 Options: ['729 cal/s', '400 cal/s', '850 cal/s', '500 cal/s', '650 cal/s', '1000 cal/s', '300 cal/s', '900
 → cal/s', '600 cal/s', '475 cal/s']
 1259 (Question 42) question: A beam of monochromatic light entering a glass window pane from the air will
 → experience a change in
 1260 Options: ['speed and wavelength', 'speed, frequency and wavelength', 'no change at all', 'frequency only',
 → 'wavelength only', 'speed, wavelength, and direction', 'speed only', 'frequency and wavelength',
 → 'speed and frequency', 'wavelength and frequency']
 1261 (Question 43) question: In what ways might international measures be employed? Are they a help or a hindrance?
 1262 Options: ['Multi-partner initiatives to support the reforms necessary for the prevention of terrorism have
 → proved ineffective. Sanctions against countries that provide a territorial base for terrorism may be
 → misplaced, having negligible effects on the terrorist organizations themselves.', 'Countries often
 → resist international measures out of concerns over sovereignty. This can limit the effectiveness of
 → these measures and hinder international cooperation against terrorism.', 'A great deal of
 → international diplomacy has involved attempts to define terrorism so that all countries could take
 → steps to eradicate it. However, these efforts have largely faltered: countries often support or
 → sympathize with dissidents who use violence against repressive governments; governments in
 → developing countries have wanted to avoid anti-colonial struggles being labelled as terrorism; and
 → countries want to retain a degree of flexibility regarding extradition and punishment of political
 → dissidents.', 'Intelligence agencies are highly centralized, organized structures that operate at a
 → maximum functional level within their domestic context. There is no reason to presume that
 → collaboration would enhance their ability to contribute to the prevention of terrorism. The multi-
 → agency approach would also broaden the scope for potential errors, miscommunication, and
 → intelligence failures.'][br/>
 1263 (Question 44) question: A 72-year-old man comes to the physician because of a 7-month history of leg weakness
 → and dry eyes and mouth. He also has had a 10.4-kg (23-lb) weight loss over the past 4 months despite
 → no change in appetite. He has smoked one and a half packs of cigarettes daily for 50 years. He drinks
 → 4 oz of alcohol daily. He has peptic ulcer disease and emphysema. Medications include cimetidine,
 → theophylline, and low-dose prednisone. Examination shows mild ptosis. He has a barrelshaped chest.
 → Breath sounds are distant. There is moderate weakness of proximal muscles of the lower extremities.
 → Reflexes are absent. He has difficulty rising from a chair. Sensory examination shows no
 → abnormalities. An x-ray shows a hyperinflated chest and a 3 x 4-cm mass in the right hilum. His
 → neurologic findings are most likely due to a lesion involving which of the following?
 1264 Options: ['Presynaptic neuromuscular junction', 'Spinal cord', 'Muscle membrane', 'Central nervous system',
 → 'Parasympathetic nervous system', 'Motor cortex', 'Postsynaptic neuromuscular junction', 'Peripheral
 → nerve', 'Sympathetic nervous system', 'Sensory nerve']
 1265 (Question 45) question: Which is the least accurate description of legal positivism?
 1266 Options: ['It perceives law as arbitrary and without any logical structure.', 'It regards morals and law as
 → inseparable.', 'It espouses the view that there is no necessary connection between morality and law
 → .', 'It regards a legal order as a closed logical system.', 'It perceives law as commands.', 'It
 → regards law as suggestions rather than commands.', 'It asserts that law is based on personal beliefs
 → and opinions.', 'It implies that legal decisions are made purely on emotional grounds.', 'It views
 → law as inherently just and moral.', 'It suggests that morality is the only basis of law.'][br/>
 1267 (Question 46) question: A 33-year-old woman comes to the physician because of a 2-day history of mild nausea,
 → increased urinary urgency and frequency, and constipation. She also has had a 4.5-kg (10-lb) weight
 → loss during the past 2 weeks and a 3-week history of vaginal bleeding. Pelvic examination shows a
 → nodular cervix with an irregular, friable posterior lip, and a rock-hard, irregular, immobile pelvic
 → mass that extends across the pelvis. Examination of biopsy specimens from the cervical and anterior
 → wall of the vagina show well-differentiated keratinizing squamous cell carcinoma. Which of the
 → following best describes the pathogenesis of this patient's disease?
 1268 Options: ['Inactivation of cellular p53', 'Insertion of viral promoters adjacent to cellular growth factor
 → genes', 'Specialized transduction', 'Transactivation of cellular growth factor genes by TAX']
 1269 (Question 47) question: Studies show that the effect of the presence of others on an individual's performance
 → is to
 1270 Options: ['Have no significant effect on the performance of tasks', 'Facilitate the performance of both easy
 → and difficult tasks', 'hinder the performance of all tasks', 'Hinder the performance of difficult

1296 → tasks', 'hinder the performance of easy tasks', 'Increase the performance of difficult tasks but
 1297 → decrease the performance of easy tasks', 'Increase the performance of novel tasks but decrease the
 1298 → performance of difficult tasks', 'facilitate the performance of easy tasks', 'Increase the
 1299 → performance of all tasks regardless of difficulty.', 'facilitate the performance of novel and
 1300 → difficult tasks']
 1300 (Question 48) question: If people expect the price of a particular product to increase in the near future
 1301 Options: ['this will increase the supply of the product.', 'this will cause the producer to decrease the
 1302 → supply of the product.', 'this will not affect the demand for the product now or later.', 'this will
 1303 → decrease the supply of the product.', 'this will increase the demand for the product.', 'this will
 1304 → not affect the demand for the product right now.', 'this will make the product obsolete.', 'this
 1305 → will decrease the demand for the product.', 'this will not affect the supply of the product.', 'this
 1306 → will cause the producer to increase the supply of the product.'])
 1306 (Question 49) question: What do the results of studies with twins suggest about the influence of heredity on
 1307 → personality?
 1307 Options: ['Heredity dictates personality completely, and the environment has no impact.', 'Identical twins
 1308 → raised apart will have completely different personalities due to environmental differences.', '
 1309 → Personality traits are inherited in a fixed pattern, similar to physical traits like eye color.', '
 1310 → Heredity determines intelligence, which in turn completely determines personality.', 'Heredity has
 1311 → no influence on personality.', 'Personalities are influenced only by peer interactions, not by
 1312 → heredity or family environment.', 'Fraternal twins have more similar personalities than identical
 1313 → twins due to shared environments.', 'Heredity provides a capacity for a wide range of behavior that
 1314 → may then be encouraged or sustained by the environment.', 'Environment solely determines personality
 1315 → traits.', 'Identical twins always have identical personalities.')]
 1315 (Question 50) question: A 64-year-old man with non-Hodgkin lymphoma comes to the physician because of a 3-week
 1316 → history of progressive numbness in his hands and feet and weakness in his legs when he stands. He
 1317 → received his third course of chemotherapy 4 weeks ago. Physical examination shows areflexia. Which of
 1318 → the following drugs is the most likely cause of these adverse effects?
 1318 Options: ['Doxorubicin', 'Fluorouracil', 'Methotrexate', 'Vincristine']
 1319 (Question 51) question: A 13-month-old child is brought to the emergency department because of urticaria,
 1320 → swelling of the lips, and difficulty breathing immediately after eating an egg. A potential risk for
 1321 → hypersensitivity reaction is posed by vaccination against which of the following illnesses?
 1321 Options: ['Influenza', 'Rotavirus', 'Human papillomavirus', 'Poliomyelitis', 'Measles', 'Meningitis', '
 1322 → Varicella', 'Tuberculosis', 'Pertussis', 'Hepatitis']
 1322 (Question 52) question: A boat travels directly upstream in a river, moving with constant but unknown speed v
 1323 → with respect to the water. At the start of this trip upstream, a bottle is dropped over the side.
 1324 → After 15 minutes the boat turns around and heads downstream. It catches up with the bottle when the
 1325 → bottle has drifted one mile downstream from the point at which it was dropped into the water. What is
 1326 → the current in the stream?
 1326 Options: ['2 miles per hour', '1 mile per hour', '2.5 miles per hour', '3 miles per hour', '5 miles per hour
 1327 → ', '4 miles per hour', '1.5 miles per hour', '0.5 miles per hour', '0.75 miles per hour', '3.5 miles
 1328 → per hour']
 1328 (Question 53) question: What is the value of the angular-momentum quantum number $\$1\$$ for a $\$t\$$ orbital?
 1329 Options: ['5', '12', '10', '7', '6', '3', '2', '16', '14', '9']
 1329 (Question 54) question: The list price of a book is \$4.95. It sells in a bookstore at a net price of \$2.95.
 1330 → What is the amount of trade discount?
 1330 Options: ['\$3.95', '\$4.00', '\$2.95', '\$2.50', '\$0.95', '\$1.00', '\$1.50', '\$1.95', '\$2.00', '\$3.00']
 1331 (Question 55) question: Sampling error refers to the
 1331 Options: ['tendency of sample statistics to be identical with each other', 'consequences of generalizing from
 1332 → a sample that is too small', 'tendency of sample statistics to always match the population
 1333 → parameters', 'consequences of generalizing from a population to a sample', 'error in the method of
 1334 → data collection in the sample', 'consequences of generalizing from a large sample to a small
 1335 → population', 'tendency of some sample statistics to differ systematically from the corresponding
 1336 → population parameters', 'tendency of sample statistics to differ from population parameters', '
 1337 → consequences of generalizing from a sample to a population']
 1337 (Question 56) question: Augustine claims that there can be no evil where there is no:
 1338 Options: ['love.', 'truth.', 'justice.', 'faith.', 'greed.', 'peace.', 'hatred.', 'compassion.', 'happiness
 1339 → .', 'good.']
 1339 (Question 57) question: The function f is given by $f(x, y, z) = [z / (x^2 + y^2)](1)$ Compute ∇f in
 1340 → cylindrical coordinates ρ , θ , z .
 1340 Options: [' $-(2z / (\rho^2))u_\rho$ ', ' $(1 / (\rho^3))u_\theta$ ', ' $(z / (\rho^3))u_z$ ', ' $(1 / (\rho^3))u_\rho + (1 / (\rho^2))u_\theta - (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_\theta + (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_z - (1 / (\rho^2))u_\theta$ ', ' $(z / (\rho^3))u_\rho - (1 / (\rho^2))u_\theta + (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_\rho + (z / (\rho^3))u_\theta - (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_\theta - (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_z - (1 / (\rho^2))u_\theta$ ', ' $(1 / (\rho^2))u_\rho - (z / (\rho^3))u_\theta - (1 / (\rho^2))u_z$ ', ' $(1 / (\rho^2))u_\theta - (z / (\rho^3))u_z - (1 / (\rho^2))u_\rho$ ']
 1341 (Question 58) question: In a lap winding dc machine number of conductors are 100 and number of parallel paths
 1342 → are 10. Find the average pitch
 1342 Options: ['150', '100', '200', '5', '10', '25', '20', '75', '50', '1']
 1343 (Question 59) question: Consider an object-oriented language in which all entities are objects. Two
 1344 → relationships arise: (1) the instance relationship, between an object and the class of which that
 1345 → object is a member, and (2) the subclass relationship, between a class and the superclass from which
 1346 → that class inherits properties. In such a language, when a message is sent to an object requesting
 1347 → execution of one of its methods (procedures), the method is located by following
 1348 Options: ['one or more instance links and then one or more subclass links', 'one or more subclass links', '
 1349 → one or more subclass links and then one or more instance links', 'one instance link and then zero or
 1349 → more subclass links', 'one or more instance links', 'one subclass link only', 'zero or more
 1349 → instance links', 'zero or more instance links and then one or more subclass links', 'one subclass
 1349 → link and then one or more instance links', 'one instance link only']
 1349 (Question 60) question: A psychologist believes that positive rewards for proper behavior are more effective
 1350 → than punishment for bad behavior in promoting good behavior in children. A scale of "proper behavior"
 1351 → is developed. 1 = the "proper behavior" rating for children receiving positive rewards, and 2 =
 1352 → the "proper behavior" rating for children receiving punishment. If $H_0: 1 = 2 = 0$, which of the
 1353 → following is the proper statement of H_A ?
 1353 Options: [' $H_A: 1 = 2 = 0$ ', 'Any of the above is an acceptable alternative to the given null.', ' $H_A:$
 1354 → $2 = 1 > 0$ ', ' $H_A: 1 = 2 \neq 0$ ', ' $H_A: 2 = 1 < 0$ ', ' $H_A: 1 = 2 \neq 0$ ', ' $H_A: 1 = 2 < 0$ ', ' $H_A:$
 1355 → $1 = 2 = 1$ ', ' $H_A: 1 = 2 > 0$ ', ' $H_A: 1 = 2 \neq 0$ ']
 1355 What main theme is present in the positive examples but absent in the negative examples? Provide a concise
 1356 → label and brief explanation.