Supplementary Materials

KC pairing rules

Below are the general guidelines employed for pairing KCs:

- The **Primary KC** must be at the Primary 3 level or higher, as multi-KC math word problems (MWPs) may be too complex for Primary 1 or Primary 2 students.
- The **Secondary KC** may be drawn from Primary 1 or Primary 2, but must be at the same or a lower grade level than the Primary KC.
- The number type associated with the Secondary KC must be the same as or simpler than that of the Primary KC. For example, if the Primary KC involves whole numbers, the Secondary KC must also involve whole numbers and cannot involve fractions or decimals.
- KCs involving operations such as comparison and ordering, conversion, rounding, evaluation, and simplification are restricted to be used only as Secondary KCs. Furthermore, these may only be paired with Primary KCs that involve the same number types.
- Division KCs are not paired with other division KCs, as it is generally difficult for two such KCs to meaningfully interact or build upon each other.

Following the application of these rules, we performed a manual review and curated 176 coherent and pedagogically compatible KC pairs for inclusion in our study.

Common reasons for not meeting evaluation criteria

We analyzed the reasons why MWPs generated by the two LLMs failed to meet the evaluation criteria. Gaining insight into these reasons can inform the design of more effective prompts with clearer instructions, thereby improving the overall quality of the generated MWPs.

Answerability Common reasons for failing to meet the *Answerability* criterion include the following:

- **Missing key information**: Essential details required to solve the problem are omitted. See Figure 1 for an example.
- **Numerical inconsistencies**: The provided numbers conflict with each other. For example, the sum of the parts exceeds the total, or items are said to be taken from a remainder when no remainder exists. See Figure 2 for an example.
- **Contradictory statements**: Logical inconsistencies in the narrative, such as stating that one quantity is less than another when it is actually greater. See Figure 3 for an example.

Primary/Secondary KC Alignment Typical causes of non-compliance with the *Primary KC Alignment* or *Secondary KC Alignment* criterion include:

- **Missing KC**: The targeted KC is not required to solve the problem. See Figure 4 for an example.

Primary KC: DECIMALS — Division — dividing decimals (up to 3 decimal places) by 10, 100, 1000 and their multiples
Secondary KC: DECIMALS — Addition — adding decimals
Topic: Recreation such as sports, games, exercises, music, movie, dancing, painting, fishing and other recreation activities
Grade: Primary 5
MWP: Wei Lin attended a swimming class where she swam a total distance of 8.250 metres on Monday and 3.750 metres on Wednesday. After her lessons each day, her coach asked her to divide the total distance she swam that week equally among 10 days, as part of a new training plan. What is the average distance, in metres, that Wei Lin should plan to swim each day, rounded to three decimal places?
Explanation: How far Wei Lin swam that week is not given.

Figure 1: Example MWP not meeting *Answerability* due to missing key information.

Primary KC: FRACTIONS — Addition — adding unlike fractions with two different denominators not exceeding 12
Secondary KC: FRACTIONS — Simplifying — expressing a fraction in its simplest form
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 4
MWP: Mrs Lim spent \$120 of her monthly expenses on groceries and \$80 on utility bills. She knew that $\frac{5}{12}$ of her expenses went to groceries and $\frac{1}{6}$ of her expenses went to utility bills. What fraction of her expenses did Mrs Lim spend on groceries and utility bills together? Give your answer in its simplest form.
Explanation: $120{:}80 \neq \frac{5}{12} : \frac{1}{6}$

Figure 2: Example MWP not meeting *Answerability* due to numerical inconsistencies.

- **Involving a more advanced related KC**: A more difficult version of the intended KC is required, such as performing the same operation on more complex numbers. See Figure 5 for an example.
- **Involving a simpler related KC**: A less complex version of the intended KC is needed, such as the same operation applied to simpler numbers. See Figure 6 for an example.
- **Involving a related KC asking for a different component**: A related KC asking for a different component is required, such as computing a total instead of number of units, or identifying a part instead of the whole. See Figure 7 for an example.

Synergy The *Synergy* criterion is violated due to the following reasons:

- **Missing Primary or Secondary KC**: One of the intended KCs is absent from the problem.

Primary KC: FRACTIONS — Division — dividing a whole number/proper fraction by a proper fraction
Secondary KC: FRACTIONS — Conversion to decimals — expressing fractions as decimals
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 6
MWP: Mrs Lee wants to divide \$120 equally among her 3 children for their weekly allowance. However, she decides to give each child only $\frac{3}{4}$ of the amount she originally planned per child, and saves the rest. (a) How much money was Mrs Lee going to give each child at first? (b) How much money does each child receive when this amount is divided by $\frac{3}{4}$? Express your answer as a decimal. (c) How much money does Mrs Lee save in total after making this change?
Explanation: Question (b) has conflicting statement. To get the answer to (b), we need to multiply the number from (a) by $\frac{3}{4}$ but it asks to divide by $\frac{3}{4}$.

Figure 3: Example MWP not meeting *Answerability* due to contradictory statements.

Primary KC: WHOLE NUMBERS — Division — dividing whole numbers up to 3 digits by 1 digit
Secondary KC: WHOLE NUMBERS — Addition — adding whole numbers
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 3
MWP: Mr. Tan received \$285 from his part-time job for cleaning houses. He wants to save this money equally over 3 months. After saving for 3 months, he decides to add \$50 to his total savings. How much money does Mr. Tan have altogether after 3 months?
Explanation: Primary KC is not required to solve the problem. Only addition is needed.

Figure 4: Example MWP not meeting *KC Alignment* due to missing KC.

- **Lack of integration between sub-questions**: The sub-questions are unrelated and do not build upon each other. See Figure 8 for an example.

Real-world Plausibility The following are common reasons for failing to meet the *Real-world Plausibility* criterion:

- **Implausible numerical values**: Quantities are unrealistically high or low for the given context–for example, running or swimming speeds exceeding world records, or prices that are excessively low or high. See Figure 9 for an example.
- **Improper monetary formatting**: Amounts of money or bills include more than two decimal places, and digits beyond the hundredths place are non-zero. This does not apply to unit pricing. See Figure 10 for an example.

Primary KC: DECIMALS — Multiplication — multiplying decimals (up to 3 decimal places) by 10, 100, 1000 and their multiples
Secondary KC: DECIMALS — Comparison and ordering — comparing and ordering decimals up to 3 decimal places
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 5
MWP: Mrs Tan received her monthly utility bill, which showed that her electricity usage for May was 3.285 kilowatt-hours (kWh) each day. The electricity company charges \$0.19 per kWh. (a) Calculate Mrs Tan's total electricity usage for May (31 days) in kWh. (b) Find the total amount Mrs Tan needs to pay for electricity used in May. (c) Mrs Tan compared her May usage of 3.285 kWh per day with her friend Mr Lee, who used 3.295 kWh per day. Who used more electricity per day and by how much?
Explanation: Question (a) requires multiplying 3.285 with 31, which is harder than the Primary KC.

Figure 5: Example MWP not meeting *KC Alignment* due to involving a more advanced related KC.

Primary KC: FRACTIONS — Multiplication — multiplying two improper fractions
Secondary KC: FRACTIONS — Addition — adding fractions
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 5
MWP: Mr Lim decided to buy 3 identical cartons of fruit juice for his family's gathering. Each carton contains $\frac{9}{4}$ litres of fruit juice. After the gathering, $\frac{5}{3}$ litres from each carton were left over. (a) What is the total amount of fruit juice in all 3 cartons before the gathering? (b) What is the total amount of fruit juice left after the gathering? (c) How much fruit juice was consumed during the gathering in total?
Explanation: For (a) and (b), a multiplication of an improper fraction and a whole number is required, which is easier than the original Primary KC.

Figure 6: Example MWP not meeting *KC Alignment* due to involving a simpler related KC.

- **Physically impossible actions or entities**: Scenarios involve partial actions or subjects that cannot exist or occur in reality. See Figure 11 for an example.
- **Unrealistic behavior**: Events or actions do not reflect typical real-world behavior—for instance, game points increasing or decreasing without a reasonable cause. See Figure 12 for an example.

Clarity Typical issues that prevent a problem from meeting the *Clarity* criterion include the following:

- **Lack of description**: Numerical values are presented

Primary KC: RATE — Finding number of unit — finding number of units given rate and total amount
Secondary KC: DECIMALS — Conversion to smaller units — converting a measurement from a larger unit in decimal form to a smaller unit
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 5
MWP: Mr. Tan checked his electricity bill and saw that he used 36.5 kilowatt-hours (kWh) of electricity in June. Each kilowatt-hour costs $0.25. How much did Mr. Tan spend on electricity in June? Give your answer in cents.
Explanation: The problem ask for the total, but the intended Primary KC asks for the number of unit.

Figure 7: Example MWP not meeting *KC Alignment* due to involving a related KC asking for a different component.

Primary KC: FRACTIONS — Subtraction — subtracting two related fractions within one whole with denominators of given fractions not exceeding 12
Secondary KC: FRACTIONS — Addition — adding fractions
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 3
MWP: Sarah spent $\frac{5}{12}$ of her pocket money on buying snacks and $\frac{1}{4}$ on buying a storybook. What fraction of her pocket money did she spend in total? How much more did she spend on snacks than on the storybook?
Explanation: the two sub-questions are not integrated.

Figure 8: Example MWP not meeting *Synergy* due to lack of integration between sub-questions.

without sufficient description, requiring the reader to infer their meaning from context. See Figure 13 for an example.

- **Ambiguity**: The problem statement can be interpreted in multiple ways, leading to confusion or multiple possible solutions. See Figure 14 for an example.

Conciseness and Relevance The *Conciseness and Relevance* criterion is not met due to the following reasons:

- **Inclusion of irrelevant information**: The problem contains extra details or numerical values that are not required to arrive at the solution. See Figure 15 for an example.
- **Redundant phrasing**: The same information is repeated unnecessarily within the problem narrative. See Figure 16 for an example.
- **Excessive repetition of the same KC**: The same knowledge component is applied three or more times to different numbers, adding unnecessary length. See Figure 17 for an example.
- **Unnecessary sub-questions**: Additional questions are included that are not essential for addressing the intended pair of KCs. See Figure 18 for an example.

Primary KC: DECIMALS — Division — dividing decimals (up to 3 decimal places) by 10, 100, 1000 and their multiples
Secondary KC: DECIMALS — Rounding — rounding decimals up to 3 decimal places to the nearest whole number, 1 decimal place and 2 decimal places
Topic: Recreation such as sports, games, exercises, music, movie, dancing, painting, fishing and other recreation activities
Grade: Primary 5
MWP: Jia Wen recorded the time she took to complete each lap while swimming for her training. She swam a total distance of 500 metres and her total time taken was 202.457 seconds. (a) What was Jia Wen's average time in seconds for each 10 metres of swimming? Express your answer as a decimal correct to 3 decimal places. (b) Round your answer in (a) to the nearest whole number, 1 decimal place, and 2 decimal places.
Explanation: Jia Wen's swimming speed is faster than word record.

Figure 9: Example MWP not meeting *Real-world Plausibility* due to implausible numerical values.

Primary KC: DECIMALS — Subtraction — subtracting decimals (up to 2 decimal places)
Secondary KC: DECIMALS — Comparison and ordering — comparing and ordering decimals up to 3 decimal places
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 4
MWP: Mrs Tan checked three of her utility bills for the month. The amounts were: $54.28, $54.105, and $53.99. (a) Arrange the amounts in order from the smallest to the largest. (b) Mrs Tan paid the highest bill last month. How much less did she pay for the lowest bill compared to the highest bill this month, rounded to 2 decimal places?
Explanation: Bill amount of the 2nd month has 3 decimal places.

Figure 10: Example MWP not meeting *Real-world Plausibility* due to improper monetary formatting.

Consistency of LLM evaluators over different runs

When using LLMs to evaluate MWPs, each MWP was assessed five times. The final label for each evaluation criterion was determined by a majority vote across these five evaluation runs. Figures 19 and 20 present the alignment between individual runs and the majority vote, measured using percentage of agreement and Cohen's kappa, for MWPs generated by GPT-4.1 and Claude Sonnet 4, respectively.

Primary KC: RATIO — Representation and concept — ratios involving decimals
Secondary KC: RATIO — Simplifying — converting a ratio involving decimals to its simplest form
Topic: Recreation such as sports, games, exercises, music, movie, dancing, painting, fishing and other recreation activities
Grade: Secondary O-level 1
MWP: During a basketball training session, Sarah made 12.5 successful free throws out of 17.5 attempts, while David made 8.4 successful free throws out of 11.2 attempts. Find the ratio of Sarah's successful free throws to David's successful free throws in its simplest form.
Explanation: One cannot have partial successful free throws.

Figure 11: Example MWP not meeting *Real-world Plausibility* due to physically impossible actions or entities.

Primary KC: WHOLE NUMBERS — Division — dividing whole numbers up to 3 digits by 1 digit with remainder.
Secondary KC: WHOLE NUMBERS — Multiplication — multiplying whole numbers
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 3
MWP: Mr Lim has $156 and wants to buy packs of apples for his children. Each pack costs $7. After buying as many packs as possible with his money, he gives 2 packs to each of his 8 neighbours. How many packs of apples does Mr Lim have left after giving to his neighbours?
Explanation: People buy items based on needs not based on how much money they have.

Figure 12: Example MWP not meeting *Real-world Plausibility* due to unrealistic behavior.

When GPT-4.1 served as the evaluator, the percentage of agreement between individual runs and the majority vote exceeded 95% for all criteria, with Cohen's kappa values above 0.68. When Claude Sonnet 4 was used as the evaluator, the percentage of agreement was above 92%, with Cohen's kappa values exceeding 0.72 across all criteria. These results indicate that, for both LLMs, individual evaluation runs are largely consistent with the majority vote.

Cosine Similarity and Euclidean Distance between MWPs within and across topics

We evaluated the similarity of MWPs both within and across topics by computing pairwise Cosine similarity and Euclidean distance based on their BGE embeddings. Figure 21 presents the mean Cosine similarity and Euclidean distance for MWPs within the same topic and across different topics. As expected, MWPs within the same topic exhibit higher similarity. Notably, the Cosine similarity within the *Recreation* topic is relatively low, suggesting a wider diversity of

Primary KC: RATIO — Finding ratio — finding the ratio of two or three given whole numbers
Secondary KC: WHOLE NUMBERS — Subtraction — subtracting whole numbers
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 6
MWP: The Tan family's monthly household expenses are $2400. They spend $800 on groceries, $600 on utilities, and the rest on other expenses. After reviewing their budget, they decided to reduce their grocery spending by $200 and their utility spending by $150. Find the ratio of their new grocery spending to their new utility spending to their new spending on other expenses.
Explanation: The problem does not mention whether the spending on other expenses changes or not.

Figure 13: Example MWP not meeting *Clarity* due to lack of description.

Primary KC: WHOLE NUMBERS — Division — dividing whole numbers up to 4 digits by 1 digit
Secondary KC: WHOLE NUMBERS — Addition — adding whole numbers
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 4
MWP: The Lim family pays a total of $3264 for their household utilities over 8 months. They pay the same amount each month. In addition to their monthly utility bill, they also pay a one-time security deposit of $156 and a one-time connection fee of $72. What is the total amount the Lim family pays for utilities including all the additional fees?
Explanation: It is not clear whether the total is over 8 months or per month. If it is over 8 months, then division is not required.

Figure 14: Example MWP not meeting *Clarity* due to ambiguity.

MWPs in this domain. Furthermore, MWPs generated by GPT and Claude for the same topic demonstrate high similarity, indicating homology across models.

Sub-clusters within each topic

We analyzed the diversity of MWPs within each topic to determine the presence of distinct sub-clusters. Figure 22 illustrates clear sub-clustering within the GPT *Services* and Claude *Recreation* topics. Further examination (Figure 23) identifies the sub-clusters in GPT *Services* as pertaining to cleaning and streaming services, while those in Claude *Recreation* correspond to sport training, piano playing, and team games. Within these subtopics, MWPs are densely concentrated, in contrast to the more sparsely and evenly distributed MWPs observed in other topics.

Primary KC: FRACTIONS — Multiplication — multiplying a proper/improper fraction and a whole number
Secondary KC: FRACTIONS — Conversion to decimals — expressing fractions as decimals
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 5
MWP: Mrs Tan uses $\frac{4}{5}$ of a packet of rice each week. If she buys 7 packets of rice, how many packets of rice will she have used after 7 weeks? Express your answer as a fraction first, then convert it to a decimal.
Explanation: Information "If she buys 7 packets of rice" is not needed for solving the problem.

Figure 15: Example MWP not meeting *Conciseness and Relevance* due to inclusion of irrelevant information.

Primary KC: DECIMALS — Division — dividing decimals (up to 2 decimal places) by a 1-digit whole number
Secondary KC: DECIMALS — Comparison and ordering — comparing and ordering decimals up to 3 decimal places
Topic: Recreation such as sports, games, exercises, music, movie, dancing, painting, fishing and other recreation activities
Grade: Primary 4
MWP: Sarah is training for a swimming competition. During her practice sessions this week, she recorded her times for swimming 400 meters. On Monday, she completed the distance in 12.48 minutes. On Tuesday, her time was 11.76 minutes. On Wednesday, she finished in 13.32 minutes. Sarah's coach told her that to find her average time per 100 meters for each day, she needs to divide each total time by 4. Help Sarah calculate her average time per 100 meters for each day, then arrange these average times from fastest to slowest.
Explanation: "finding average time per 100 meters for each day" is repeated twice.

Figure 16: Example MWP not meeting *Conciseness and Relevance* due to redundant phrasing.

MWP generation prompts to the two LLMs

Figures 24 and 25 present the system prompt and user prompt, respectively, used with GPT for MWP generation. Figure 26 displays the user prompt employed for MWP generation with Claude Sonnet 4, which does not utilize a system prompt.

MWP evaluation prompts to the two LLMs

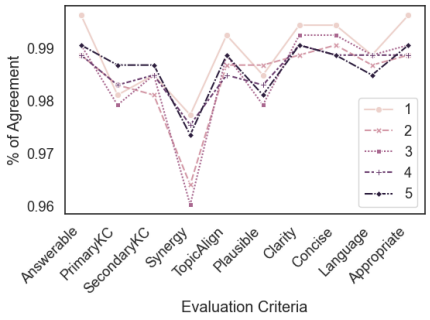
Figures 27 and 28 show the system prompt and user prompt, respectively, provided to GPT for MWP evaluation. Figure 29 depicts the user prompt used for MWP evaluation with Claude Sonnet 4.

Primary KC: DECIMALS — Addition — adding decimals (up to 2 decimal places)
Secondary KC: DECIMALS — Comparison and ordering — comparing and ordering decimals up to 3 decimal places
Topic: Household finance such as income, utility bills, money, interest, savings, instalment, mortgage, financial planning etc.
Grade: Primary 4
MWP: The Tan family is comparing their monthly utility bills for the past three months to plan their budget better. In January, their electricity bill was \$45.60 and their water bill was \$23.40. In February, their electricity bill was \$48.75 and their water bill was \$21.85. In March, their electricity bill was \$44.20 and their water bill was \$25.30. (a) What was the total utility bill for each month? (b) Arrange the three months in order from the lowest total utility bill to the highest total utility bill.
Explanation: Primary KC needs to be applied three times. The problem can actually give the total utility bill in Feb and Mar for the comparision without the excessive computation.

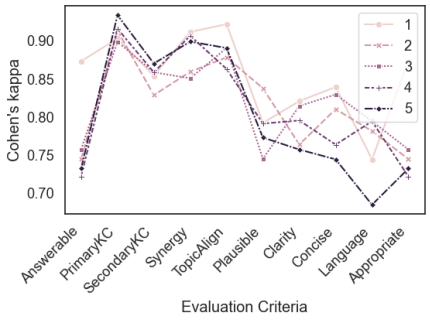
Figure 17: Example MWP not meeting *Conciseness and Relevance* due to excessive repetition of the same KC.

Primary KC: MATRICES — Subtraction — subtraction of matrices
Secondary KC: MATRICES — Addition — addition of matrices
Topic: Recreation such as sports, games, exercises, music, movie, dancing, painting, fishing and other recreation activities
Grade: Secondary O-level 3/4
MWP: A gaming club tracks the scores of their members across three different video games over two consecutive weeks. The scores are recorded in matrices where each row represents a different game and each column represents a different player.
Week 1 scores matrix A: $A = \begin{pmatrix} 85 & 92 & 78 \\ 76 & 88 & 95 \\ 90 & 82 & 87 \end{pmatrix}$
Week 2 scores matrix B: $B = \begin{pmatrix} 88 & 89 & 82 \\ 80 & 85 & 92 \\ 87 & 86 & 90 \end{pmatrix}$.
The club wants to analyze the performance changes and calculate some statistics. (a) Find the matrix C that represents the change in scores from Week 1 to Week 2 (Week 2 scores minus Week 1 scores). (b) Calculate the total scores matrix D for both weeks combined. (c) If the club gives bonus points represented by matrix $E = \begin{pmatrix} 5 & 3 & 4 \\ 2 & 6 & 1 \\ 3 & 4 & 5 \end{pmatrix}$ to the total scores, what is the final scores matrix F?
Explanation: Question (c) is not needed to cover the two KCs.

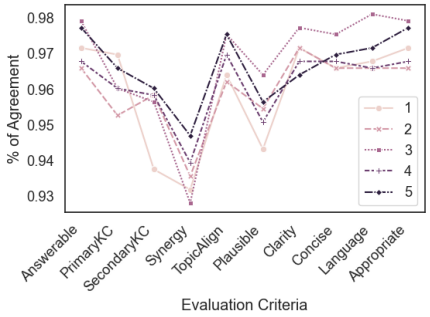
Figure 18: Example MWP not meeting *Conciseness and Relevance* due to unnecessary sub-questions.



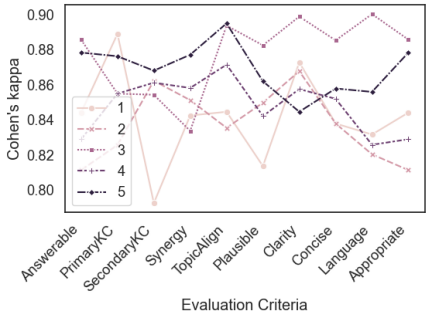
(a) GPT evaluator, % of Agreement



(b) GPT evaluator, Cohen's kappa

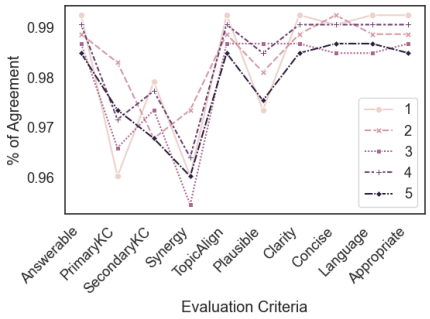


(c) Claude evaluator, % of Agreement

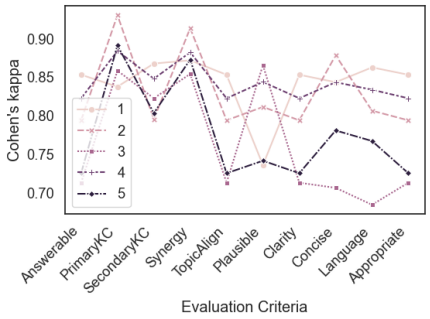


(d) Claude evaluator, Cohen's kappa

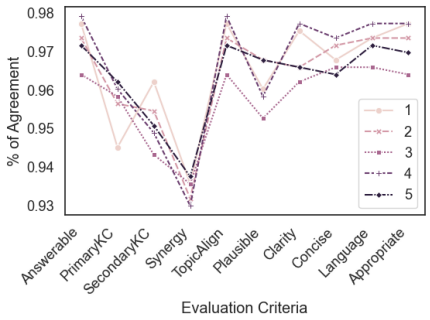
Figure 19: Alignment of 5 runs with majority voting on MWPs generated by GPT-4.1



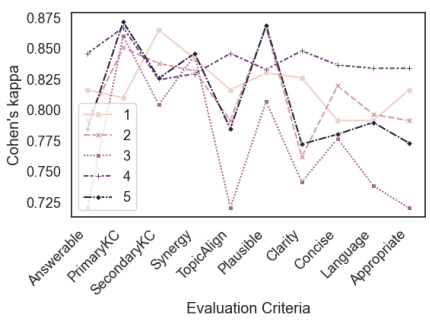
(a) GPT evaluator, % of Agreement



(b) GPT evaluator, Cohen's kappa

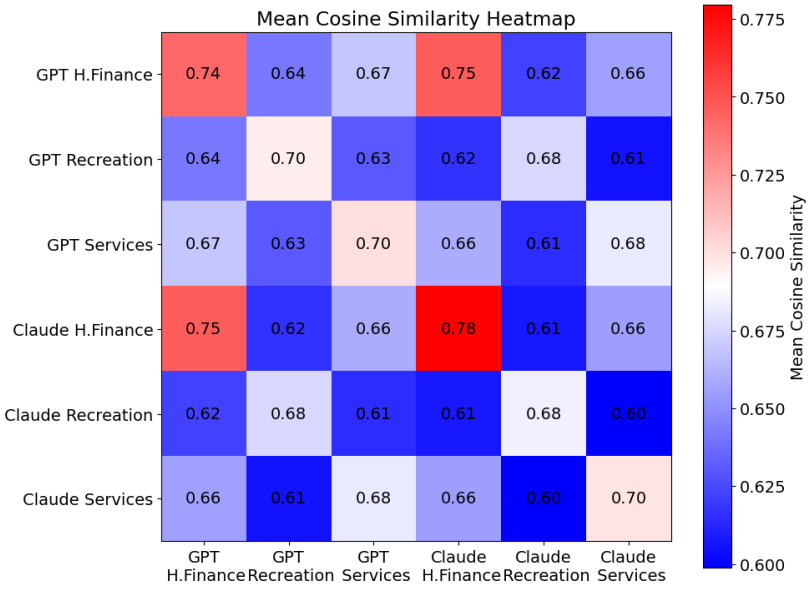


(c) Claude evaluator, % of Agreement

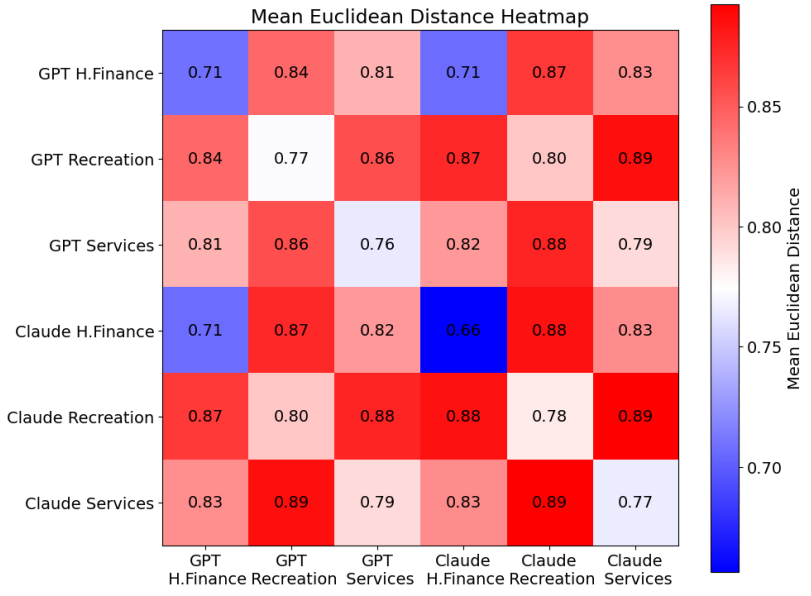


(d) Claude evaluator, Cohen's kappa

Figure 20: Alignment of 5 runs with majority voting on MWPs generated by Claude Sonnect 4

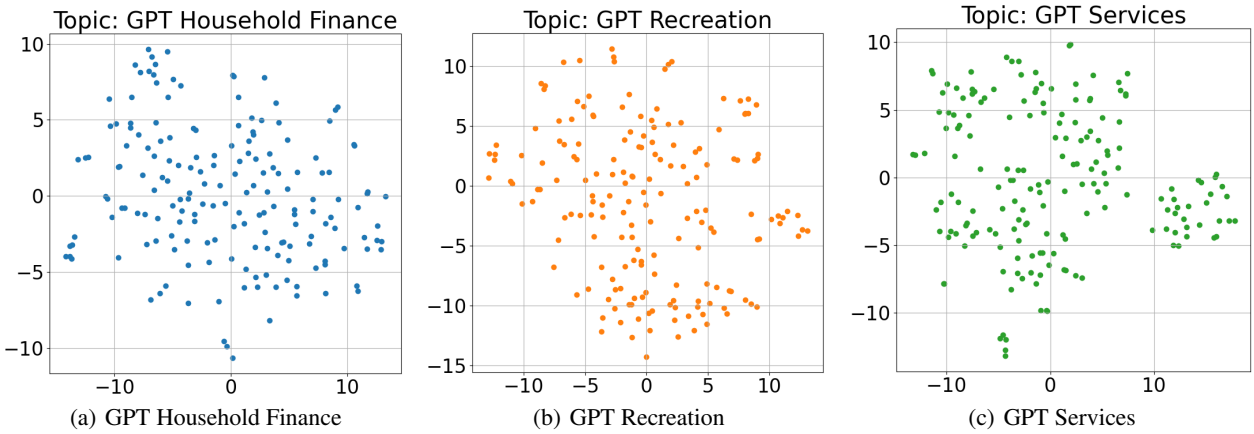


(a) Cosine similarity



(b) Euclidean distance

Figure 21: Mean Cosine Similarity and Euclidean Distance between MWPs within and across topics



(a) GPT Household Finance (b) GPT Recreation (c) GPT Services

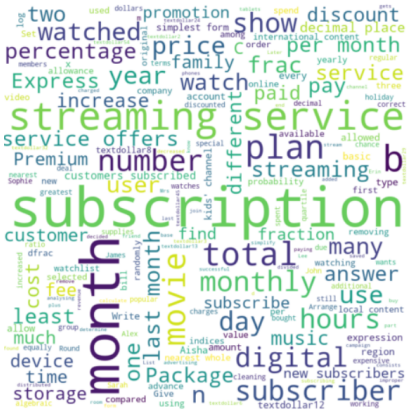
(d) Claude Household Finance (e) Claude Recreation (f) Claude Services

Figure 22: Sub-clusters within each topic

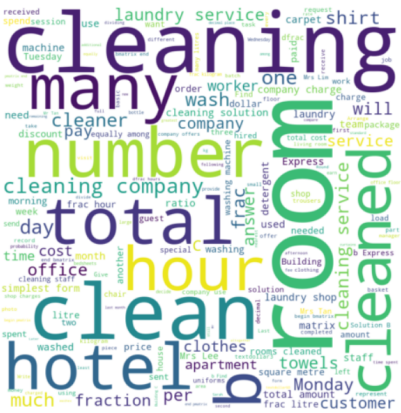
(a) Claude Recreation Sub-cluster 1

(b) Claude Recreation Sub-cluster 2

(c) Claude Recreation Sub-cluster 3



(d) GPT Services Sub-cluster 1



(e) GPT Services Sub-cluster 2

Figure 23: Sub-clusters word cloud for GPT Services and Claude Recreation

Role and Objective
You are an educational content generator specialized in **CREATING MATH WORD PROBBLEM**.
Your task is to write a pedagogically sound and realistic **MATH WORD PROBLEM**, along with its **SOLUTION**. The word problem must **FOCUS ON THE GIVEN PRIMARY KNOWLEDGE COMPONENT** and **THE GIVEN SECONDARY KNOWLEDGE COMPONENT**, and **FALL UNDER THE GIVEN TOPIC**.

Requirements
1. Answerability: The word problem must be answerable with NO missing information, NO conflicting information and NO illogical relationship.
2. Primary Knowledge Component Alignment: The word problem MUST be on the intended Primary Knowledge Component, that is, the Primary Knowledge Component is needed for solving the word problem. The Primary Knowledge Component decides the main idea and difficulty level of the word problem and there should be no other knowledge component in the word problem that is harder than it.
3. Secondary Knowledge Component Alignment: The word problem MUST also target the intended Secondary Knowledge Component, that is, the Secondary Knowledge Component is also needed for solving the word problem. The Secondary Knowledge Component adds complexity to the word problem.
4. Synergy: Primary Knowledge Component and Secondary Knowledge Component need to be used in tandem to solve the word problem. If there are multiple questions in the word problem, the questions need to be connected with each other such that the result of an early question should be used as known information to a later question in the word problem.
5. Topic Alignment: The context of the word problem MUST belong to the given Topic.
6. Grade Alignment: The word problem MUST be suitable for the given Student Grade. It should not require any Knowledge Components above the given Grade.
7. Real-World Plausibility: The word problem MUST describe a realistic scenario mirroring how people typically act and operate in reality.
8. Clarity: The language used in the word problem must be easy to understand and free from ambiguity.
9. Conciseness and Relevance: There is no irrelevant information in the word problem that is not useful for solving the problem. All given numbers must be necessary for solving the word problem.
10. Language Quality: The word problem uses correct grammar, spelling, and vocabulary.
11. Content Appropriateness: The word problem is respectful, age-appropriate, and free from offensive content.
12. Format: The format of the word problem and solution must be **IN LATEX CODE** where **ALL MATH EXPRESSIONS MUST BE IN MATH MODE $ $** and **ALL NORMAL MONEY DOLLAR SIGNS MUST BE EXPRESSED AS \textdollar<number>**, for example: \textdollar7 for $7.

Input
1. Primary Knowledge Component.
2. Secondary Knowledge Component.
3. Topic.
4. Student Grade.

Output
Please response strictly in this dictionary format: {"word_problem": "Your generated word problem here", "solution": "Your generated solution here"}.

Figure 24: System prompt to GPT-4.1 for MWP generation.

Generate a Math Word Problem based on the following information:
1. The primary Knowledge Component is {kc1}
2. The secondary Knowledge Component is {kc2}
3. The topic is {topic}
4. The student grade is {grade}

Figure 25: User prompt to GPT-4.1 for MWP generation.

<Role and Objective>
You are an educational content generator specialized in creating math word problems.
Your task is to write a pedagogically sound and realistic math word problem, along with its solution. The word problem must focus on the given primary knowledge component and the given second knowledge component,and fall under the given topic.

<Requirements>
1. Answerability: The word problem must be answerable with NO missing information, NO conflicting information and NO illogical relationship.
2. Primary Knowledge Component Alignment: The word problem MUST be on the intended Primary Knowledge Component, that is, the Primary Knowledge Component is needed for solving the word problem. The Primary Knowledge Component decides the main idea and difficulty level of the word problem and there should be no other knowledge component in the word problem that is harder than it.
3. Secondary Knowledge Component Alignment: The word problem MUST also target the intended Secondary Knowledge Component, that is, the Secondary Knowledge Component is also needed for solving the word problem. The Secondary Knowledge Component adds complexity to the word problem.
4. Synergy: Primary Knowledge Component and Secondary Knowledge Component need to be used in tandem to solve the word problem. If there are multiple questions in the word problem, the questions need to be connected with each other such that the result of an early question should be used as known information to a later question in the word problem.
5. Topic Alignment: The context of the word problem MUST belong to the given Topic.
6. Grade Alignment: The word problem MUST be suitable for the given Student Grade. It should not require any Knowledge Components above the given Grade.
7. Real-World Plausibility: The word problem MUST describe a realistic scenario mirroring how people typically act and operate in reality.
8. Clarity: The language used in the word problem must be easy to understand and free from ambiguity.
9. Conciseness and Relevance: There is no irrelevant information in the word problem that is not useful for solving the problem. All given numbers must be necessary for solving the word problem.
10. Language Quality: The word problem uses correct grammar, spelling, and vocabulary.
11. Content Appropriateness: The word problem is respectful, age-appropriate, and free from offensive content.
12. Format: The format of the word problem and solution must be in latex code where all math expressions must be in math mode $ $ and all normal money dollar signs must be expressed as \textdollar<number>, for example: \textdollar7 for $7.

<Input>
Generate a Math Word Problem based on the following inputs:
1. The primary Knowledge Component is {kc1}
2. The secondary Knowledge Component is {kc2}
3. The topic is {topic}
4. The student grade is {grade}

<Output>
Please output your response in a valid JSON object with the following keys: "word_problem" (str, your generated word problem) and "solution" (str, your generated solution).

Figure 26: User prompt to Claude Sonnect 4 for MWP generation.

Role and Objective
You are a strict educational content evaluator specialized in **EVALUATE THE QUALITY OF MATH WORD PROBBLEMS **. Your goal is to evaluate a **GIVEN MATH WORD PROBLEM** as strictly as possible based on the given **EVALUATION CRITERIA**. The math word problem is created using an intended Primary Knowledge Component, an intended Secondary Knowledge Component, an intended Topic and an intended Grade.

Evaluation Criteria
1. Answerability: The word problem must be answerable with NO missing information, NO conflicting information and NO illogical relationship.
2. Primary Knowledge Component Alignment: The word problem MUST be on the intended Primary Knowledge Component, that is, the Primary Knowledge Component is needed for solving the word problem. The Primary Knowledge Component decides the main idea and difficulty level of the word problem and there should be no other knowledge component in the word problem that is harder than it.
3. Secondary Knowledge Component Alignment: The word problem MUST also target the intended Secondary Knowledge Component, that is, the Secondary Knowledge Component is also needed for solving the word problem. The Secondary Knowledge Component adds complexity to the word problem.
4. Synergy: Primary Knowledge Component and Secondary Knowledge Component need to be used in tandem to solve the word problem. If there are multiple questions in the word problem, the questions need to be connected with each other such that the result of an early question should be used as known information to a later question in the word problem.
5. Topic Alignment: The context of the word problem MUST belong to the given Topic.
6. Grade Alignment: The word problem MUST be suitable for the given Student Grade. It should not require any Knowledge Components above the given Grade.
7. Real-World Plausibility: The word problem MUST describe a realistic scenario mirroring how people typically act and operate in reality.
8. Clarity: The language used in the word problem must be easy to understand and free from ambiguity.
9. Conciseness and Relevance: There is no irrelevant information in the word problem that is not useful for solving the problem. All given numbers must be necessary for solving the word problem.
10. Language Quality: The word problem uses correct grammar, spelling, and vocabulary.
11. Content Appropriateness: The word problem is respectful, age-appropriate, and free from offensive content.

Instructions
1. Evaluate the given math word problem as strictly as possible based on the given evaluation criteria.
2. On each evaluation dimension, if you think the given **MATH WORD PROBLEM** fully satisfies the criteria on the dimension, rate 1, otherwise rate 0.
3. Noted that the **MATH WORD PROBLEM** is written **IN LATEX**, ignore all the LaTeX error.
4. Provide explanations for **ALL** criteria you gave 0, using alpha-numerical characters and +,-, *, / only.

Given Inputs
1. Generated Math Word Problem
2. Intended Primary Knowledge Component.
3. Intended Secondary Knowledge Component.
4. Intended Topic.
5. Intended Student Grade.

Output
Please provide your response strictly in the JSON format below (1 for fully satisfying the given criteria, otherwise 0):
{
”answerability”: 0/1,
”primary_kc_alignment”: 0/1,
”secondary_kc_alignment”: 0/1,
”synergy”: 0/1,
”topic_alignment”: 0/1,
”grade_alignment”: 0/1,
”real_world_feasibility”: 0/1,
”clarity”: 0/1,
”conciseness_and_relevance”: 0/1,
”language_quality”: 0/1,
”content_appropriateness”: 0/1,
”explanation”: explanations for any criteria you gave a negative rating of 0, leave blank otherwise
}

Figure 27: System prompt to GPT-4.1 for MWP evaluation.

Evaluate a Math Word Problem based on the following information:
1. The Generated Math Word Problem is {word_problem}
2. The Intended Primary Knowledge Component is {kc1}
3. The Intended Secondary Knowledge Component is {kc2}
4. The Intended Topic is {topic}
5. The Intended Student Grade is {grade}

Figure 28: User prompt to GPT-4.1 for MWP evaluation.

Role and Objective:
You are a strict educational content evaluator specialized in **EVALUATE THE QUALITY OF MATH WORD PROBBLEMS**. Your goal is to evaluate a **GIVEN MATH WORD PROBLEM** as strictly as possible based on the given **EVALUATION CRITERIA**. The math word problem is created using an intended Primary Knowledge Component, an intended Secondary Knowledge Component, an intended Topic and an intended Grade.

Evaluation Criteria:
1. Answerability: The word problem must be answerable with NO missing information, NO conflicting information and NO illogical relationship.
2. Primary Knowledge Component Alignment: The word problem MUST be on the intended Primary Knowledge Component, that is, the Primary Knowledge Component is needed for solving the word problem. The Primary Knowledge Component decides the main idea and difficulty level of the word problem and there should be no other knowledge component in the word problem that is harder than it.
3. Secondary Knowledge Component Alignment: The word problem MUST also target the intended Secondary Knowledge Component, that is, the Secondary Knowledge Component is also needed for solving the word problem. The Secondary Knowledge Component adds complexity to the word problem.
4. Synergy: Primary Knowledge Component and Secondary Knowledge Component need to be used in tandem to solve the word problem. If there are multiple questions in the word problem, the questions need to be connected with each other such that the result of an early question should be used as known information to a later question in the word problem.
5. Topic Alignment: The context of the word problem MUST belong to the given Topic.
6. Grade Alignment: The word problem MUST be suitable for the given Student Grade. It should not require any Knowledge Components above the given Grade.
7. Real-World Plausibility: The word problem MUST describe a realistic scenario mirroring how people typically act and operate in reality.
8. Clarity: The language used in the word problem must be easy to understand and free from ambiguity.
9. Conciseness and Relevance: There is no irrelevant information in the word problem that is not useful for solving the problem. All given numbers must be necessary for solving the word problem.
10. Language Quality: The word problem uses correct grammar, spelling, and vocabulary.
11. Content Appropriateness: The word problem is respectful, age-appropriate, and free from offensive content.

Instructions:
1. Evaluate the given math word problem as strictly as possible based on the given evaluation criteria.
2. On each evaluation dimension, if you think the **GIVEN MATH WORD PROBLEM** fully satisfies the criteria on the dimension, rate 1, otherwise rate 0.
3. Noted that the **MATH WORD PROBLEM** is written **IN LATEX**, ignore all the LaTeX error.
4. Provide explanations for **ALL** criteria you gave 0, using alpha-numerical characters and +,-,*,/ only.

Use the inputs below and output your evaluation (1 for fully satisfying the given criteria, otherwise 0) in a valid JSON object with keys: "answerability" (binary, 0/1), "primary_kc_alignment" (binary, 0/1), "secondary_kc_alignment" (binary, 0/1), "synergy" (binary, 0/1), "topic_alignment" (binary, 0/1), "grade_alignment" (binary, 0/1), "real_world_feasibility" (binary, 0/1), "clarity" (binary, 0/1), "conciseness_and_relevance" (binary, 0/1), "clarity" (binary, 0/1), "language_quality" (binary, 0/1), "content_appropriateness" (binary, 0/1), and "explanation" (str, explanations for any criteria you gave a negative rating of 0, leave blank otherwise).

Evaluate the given Math Word Problem below which is created using the intended knowledge components, topic and grade below:
1. The Given Math Word Problem is {word_problem}
2. The Intended Primary Knowledge Component is {kc1}
3. The Intended Secondary Knowledge Component is {kc2}
4. The Intended Topic is {topic}
5. The Intended Student Grade is {grade}

Figure 29: User prompt to Claude Sonnect 4 for MWP evaluation.