# Recommendation Misconception Dataset

Author: Nancy Otero

Date: November 2023

This is a dataset of research-based recommended practices to help learners overcome math misconceptions. The goal of the research-based recommendations is not to give the answer to a type of math problem or exclusively guide learners through an algorithmic procedure. They are pedagogical practices on teaching mathematical concepts and procedures related to specific misconceptions or pedagogical practices on how to teach those mathematical concepts and procedures without creating a misconception.

The dataset contains four recommendations per misconception. The chosen misconceptions are based on the [previous dataset](https://github.com/creature-ai/math-misconceptions). Each recommendation has two parts: 1) A practical description of the pedagogical approach required to overcome that misconception, and 2) A mathematical problem to which the pedagogical recommendation can be applied.

To explore the uses of the dataset with LLMs, an experiment was done where three agents were asked which of the four recommendations was the most appropriate for an error in a math problem given a misconception. Two agents were middle school math teachers, and the other was GPT4.

GPT4 and Teacher1:

* Percentage of Agreement: 90.91%
* Cohen's Kappa: 0.875

GPT4 and Teacher2:

* Percentage of Agreement: 47.27%
* Cohen's Kappa: 0.288

Teacher1 and Teacher2:

* Percentage of Agreement: 43.64%
* Cohen's Kappa: 0.242

Possible applications to the EdTech ecosystem include:

* Creation of hints and feedback for learners
* Training of teachers in possible recommendations for misconceptions
* Generation of synthetic data that can serve as simulated teachers

## 

## Dataset

An in-progress repo in Github is available for the LEVI ecosystem. The dataset includes per misconception:

### 

### **Meta Data**

* Misconception description
* ID

### **Data**

* **Four research-based recommendations** to either overcome that misconception or teach the mathematical concepts related to the misconception without creating a misconception. Each recommendation includes:
  + Recommendation description- this is a practical pedagogical approach.
  + Paper reference for the recommendation description with a link to the paper
  + Example- a mathematical problem where the recommendation description can be applied
  + Paper reference for the example with a link to the paper

Here are two examples of the Misconception, the Recommendation Description and the Recommendation Example. In most cases, the paper from the recommendation and the paper for the example of the recommendation are the same. However, the recommendation was found without a clear example in a few cases.

| **Misconception** | **ID** | **Recommendation Description 1** | **Paper** | **Recommendation Example 1** | **Paper** |
| --- | --- | --- | --- | --- | --- |

| students show difficulty depicting key aspects and relationships in patterns | MaE36 | Have students work with picture patterns to create their own rules to represent the pattern and then move forward towards non-pictorial (more abstract) generalizations. | Bush, 2001  p. 89 | Example: The Matchsticks problem  Construct matchsticks squares, using an appropriate number of matchsticks, to make 1, 2, 3, 4, ... squares, and generalize the pattern.  Examples of learners initial rules of the pattern:  When there are 3 squares, there are 10 matches  \_ \_ \_  |\_|\_|\_|  When there are 7 squares, there are 22 matches  \_ \_ \_ \_ \_ \_ \_  |\_|\_|\_|\_|\_|\_|\_| | Healy, L., & Hoyles, C. (1999)  p. 8 |
| --- | --- | --- | --- | --- | --- |

*Example of the parts of a recommendation without papers*

| Students don’t recognize the inconsistency between the positive/negative slope of the line and the negative/positive slope in the equation when incorrectly calculating or plotting | MaE41 | Make students create spreadsheets to depict their mental representations of the information when working with functions with negative slopes, and begin to describe the information in terms of domain, range, independent and dependent variables, and transformations | Davidenko, S (1997)  p. 7 | Example: The Weather Balloon Problem  Situation. Summer weather in Maryland and Pennsylvania brings heavy clouds and thunderstorms on many late afternoons. As warm, moist air rises, it cools. When the air has cooled to the condensation temperature, it forms water drops. These data were recorded by a weather balloon sent up on a warm day.  Data  Altitude in meters | Temperature in degrees centigrade  ------------------------------------------------------  0 | 32  500 | 27  1000 | 23  1500 | 18  2000 | 14.5  2500 | 9  3000 | 3.5  3500 | -3  1. Use a function-fitting program to find a linear function that describes the data well. Record the rule relating temperature, t(a), to altitude, a, rounding the coefficient and constant term appropriately  2. Explain what the slope and constant term reveal about the temperature as it is related to altitude.  3. Look at a plot of your data and the fitted function to see how well the rule matches the experimental data. Can you see any reason that the altitude and temperature data are not exactly linear? How well does the fitted function represent a reasonable range of values for the altitude? | Kilpatrick, J., Swafford, J. O., & Findell, B. (2001)  p. 291 |
| --- | --- | --- | --- | --- | --- |

*Example of the parts of a recommendation without papers*

## 

## Testing

The rationale is to create datasets that can be used to change the instructive behavior of LLMs. LLMs have been criticized as math tutors because of their inclination to either give the answers to the learners' questions instead of supporting them in learning how to answer them themselves or because of their focus on the procedural component of the answer instead of the mathematical understanding. The data that most LLMs are trained with doesn't include the best learning practices; rather, it contains interactions where someone asks a question and someone else gives the answer. This dataset is a collection of research-based best practices.

To explore the uses of the dataset with LLMs, a test was done where three agents were asked which of the four recommendations was the most appropriate for an error in a math problem. Two agents were middle school math teachers, and the other was GPT4.

The three agents were tested to select the best recommendation from the four recommendations. The information given to the agents was:

* A math question and an answer with a particular misconception
* A description of the misconception

The previous information was retrieved from the Misconception Dataset. The math question is the first example of the misconception.

The prompt was:

Hi, you are a middle school math tutor. Your students have a misconception.

Which of these recommendations will be the most useful?

Delimit your analysis to the given misconception, recommendations and their examples. Please give your final answer by just writing the number of the recommendation.

Information:

Misconception: {ex['Misconception']} \n

Example Misconception: {ex['Example1']} \n

Recommendation1: {ex['AI rec 1']} \n

Example Recommendation1: {ex['Ex 1']} \n

Recommendation2: {ex['AI rec 2']} \n

Example Recommendation2: {ex['Ex 2']} \n

Recommendation3: {ex['AI rec 3']} \n

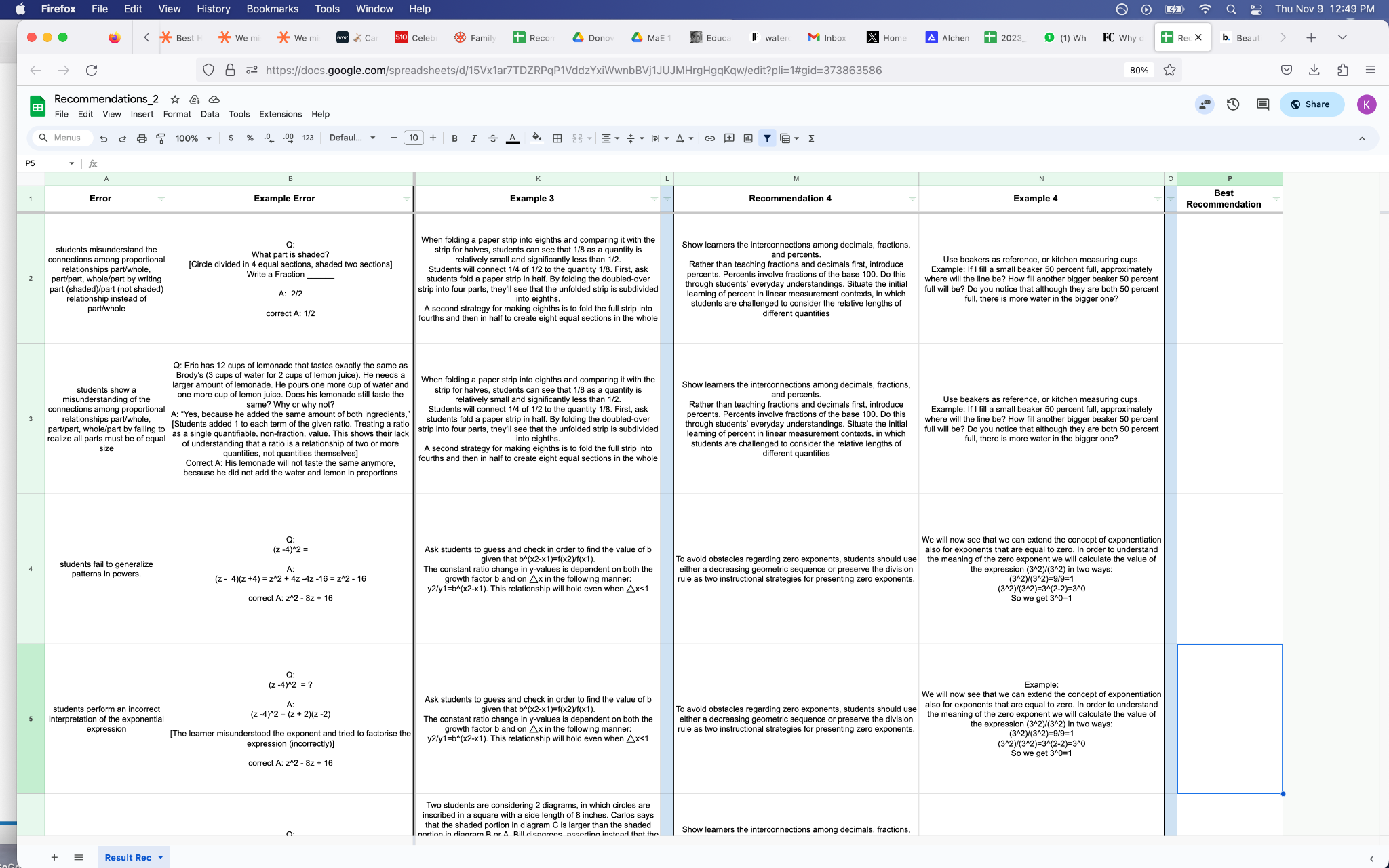
Example Recommendation3: {ex['Ex 3']} \n

Recommendation4: {ex['AI rec 4']} \n

Example Recommendation4: {ex['Ex 4']} \n

### **Comparison**

The same prompt was used with two middle school math teachers. The teachers were given a spreadsheet with the information where they wrote the number of recommendations they found best suited for the misconception example. It took them about two hours to finish the task. They were compensated with $40.



*Image of the document given to the teachers. Document*

### Results

**GPT4 and Teacher1:**

* + Percentage of Agreement: 90.91%
  + Cohen's Kappa: 0.875

**GPT4 and Teacher2:**

* + Percentage of Agreement: 47.27%
  + Cohen's Kappa: 0.288

**Teacher1 and Teacher2:**

* + Percentage of Agreement: 43.64%
  + Cohen's Kappa: 0.242