

Evaluation Instructions

We will have three files to evaluate, each for a different type of prompting strategy:

- **Zero-shot**
- **Zero-shot with an example of how to construct the question**
- **Zero-shot with a description of the construct and how to use it**

Question distribution by level:

- A1 – 3
- A2 – 6
- B1 – 6
- B2 – 6
- C1 – 3

Note! Since we are testing how well the given prompting strategy aligns with the specified parameters (number of questions included), it is possible that some prompts will generate fewer questions than expected. If a topic was skipped during prompting, I will include it in our evaluation, so you do not need to lower your score because of it.

These are all the evaluation questions I used. Please answer all of them (except the 7th one!) with **0** or **1** in the cell for the given question.

These are all the topics that were given to each model. I will also attach the updated JSON file.