

Ariel Blobstein
Ben-Gurion University

Daniel Izmaylov
Ben-Gurion University

Tal Yifat
Data Innovation Lab

Michal Levy
Tel Aviv University

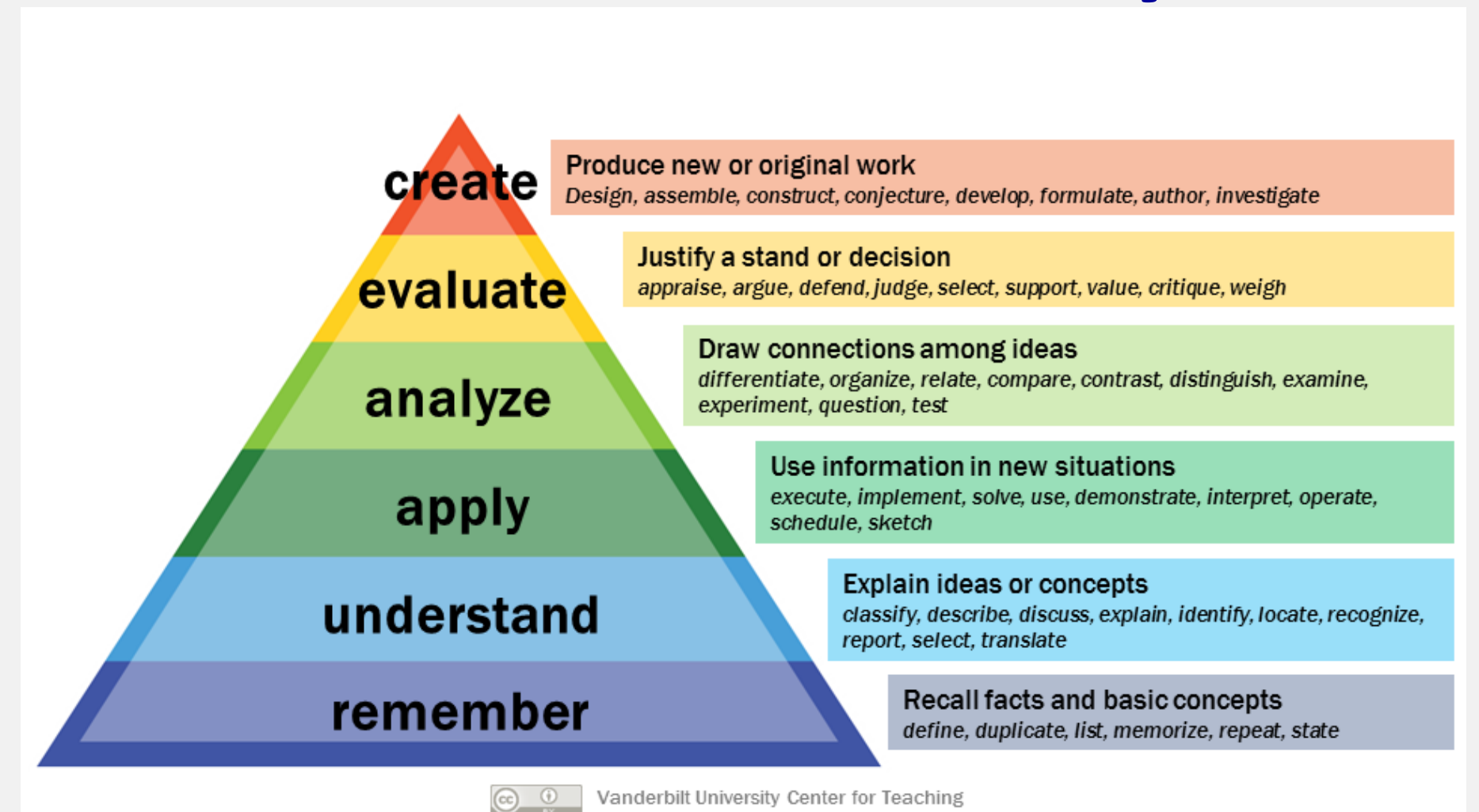
Avi Segal
Ben-Gurion University



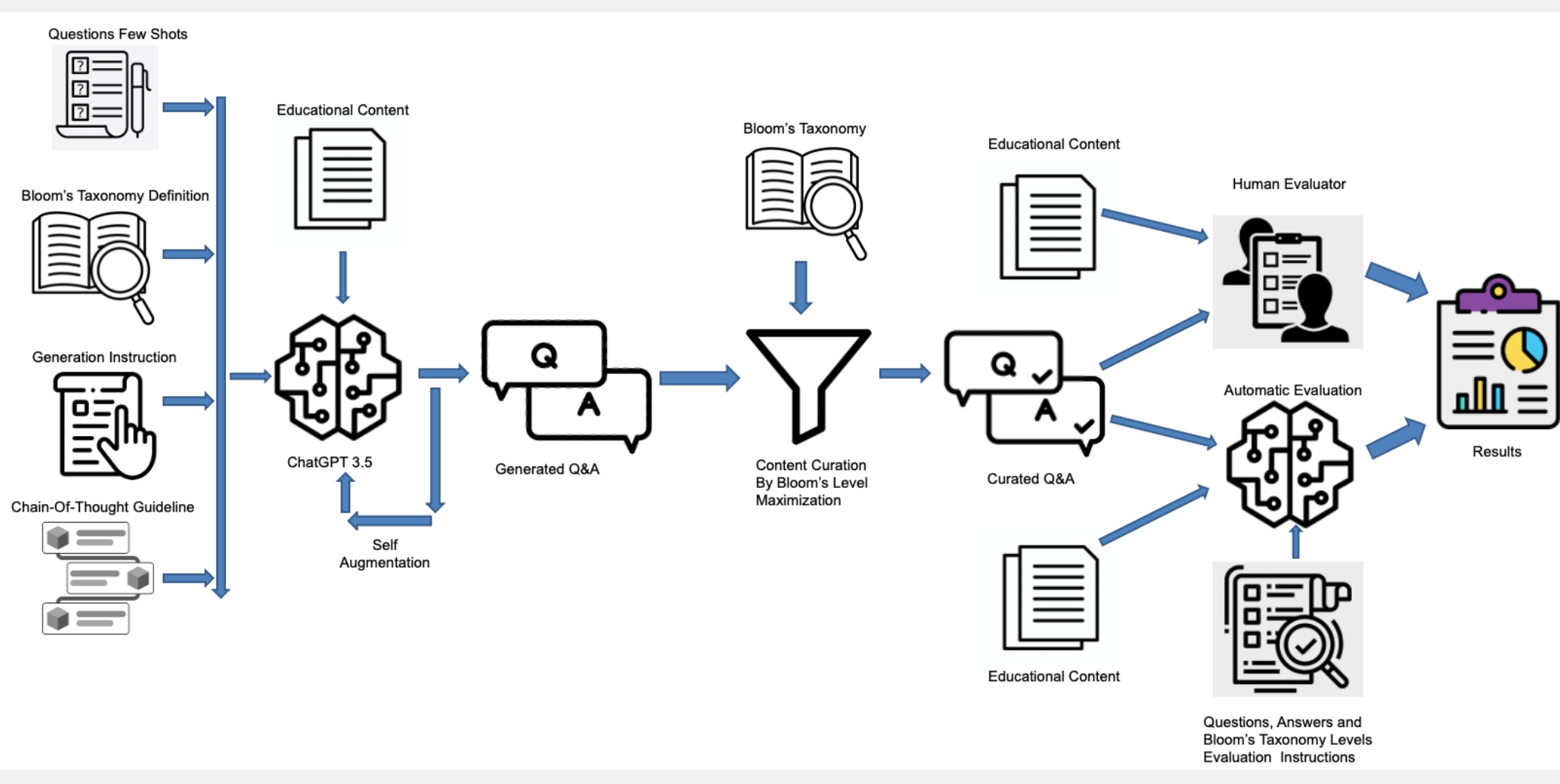
1 Background

- Automatically creating high quality Q&A for educational material continues to be a challenge
- Past attempts ignored cognitive and difficulty levels
- Bloom's taxonomy: a framework for categorizing cognitive skills for learning activities in education
- Generative AI holds pivotal potential in this domain
- Angel: Gen AI for Q&A creation, curation and evaluation
- Harnessing few-shot learning and Bloom's Taxonomy

2 Bloom's Taxonomy



3 Angel's Approach



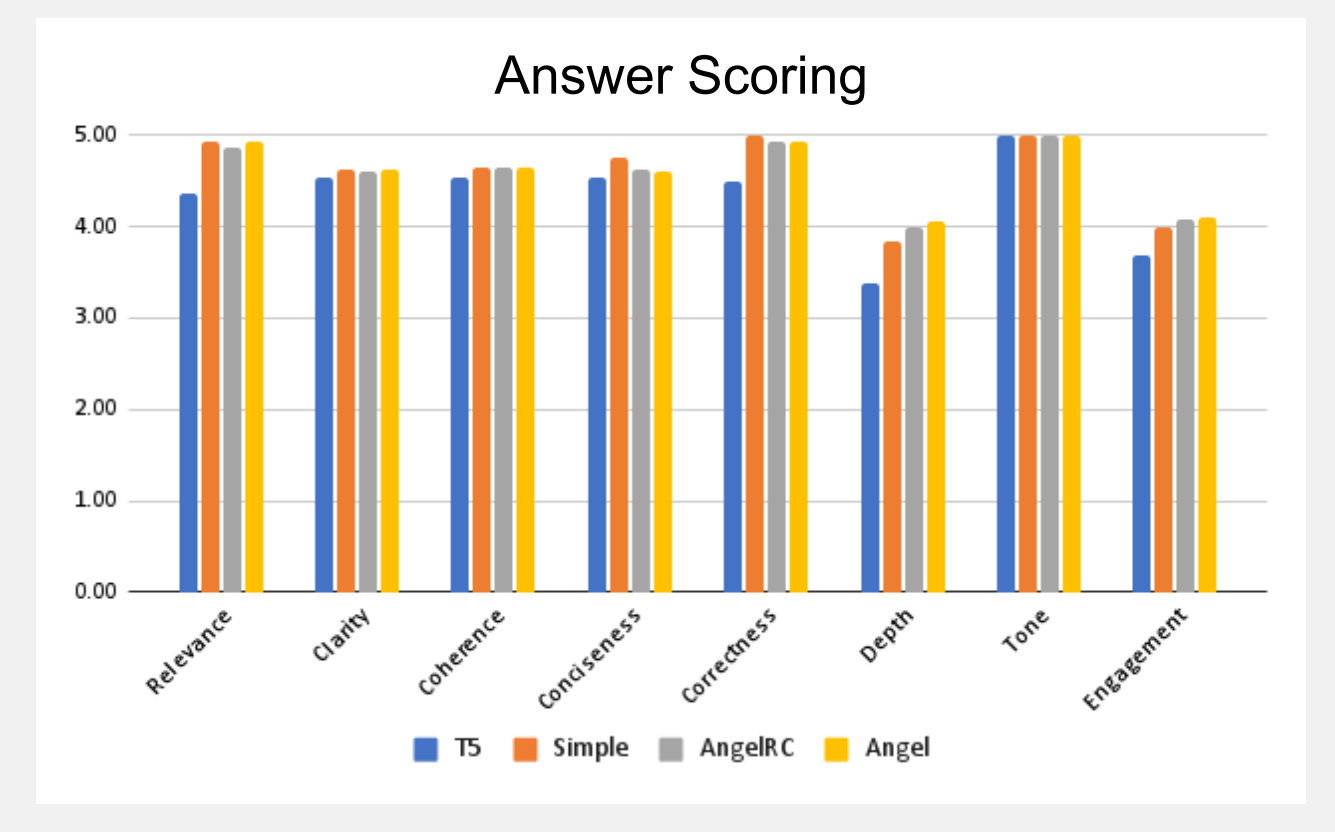
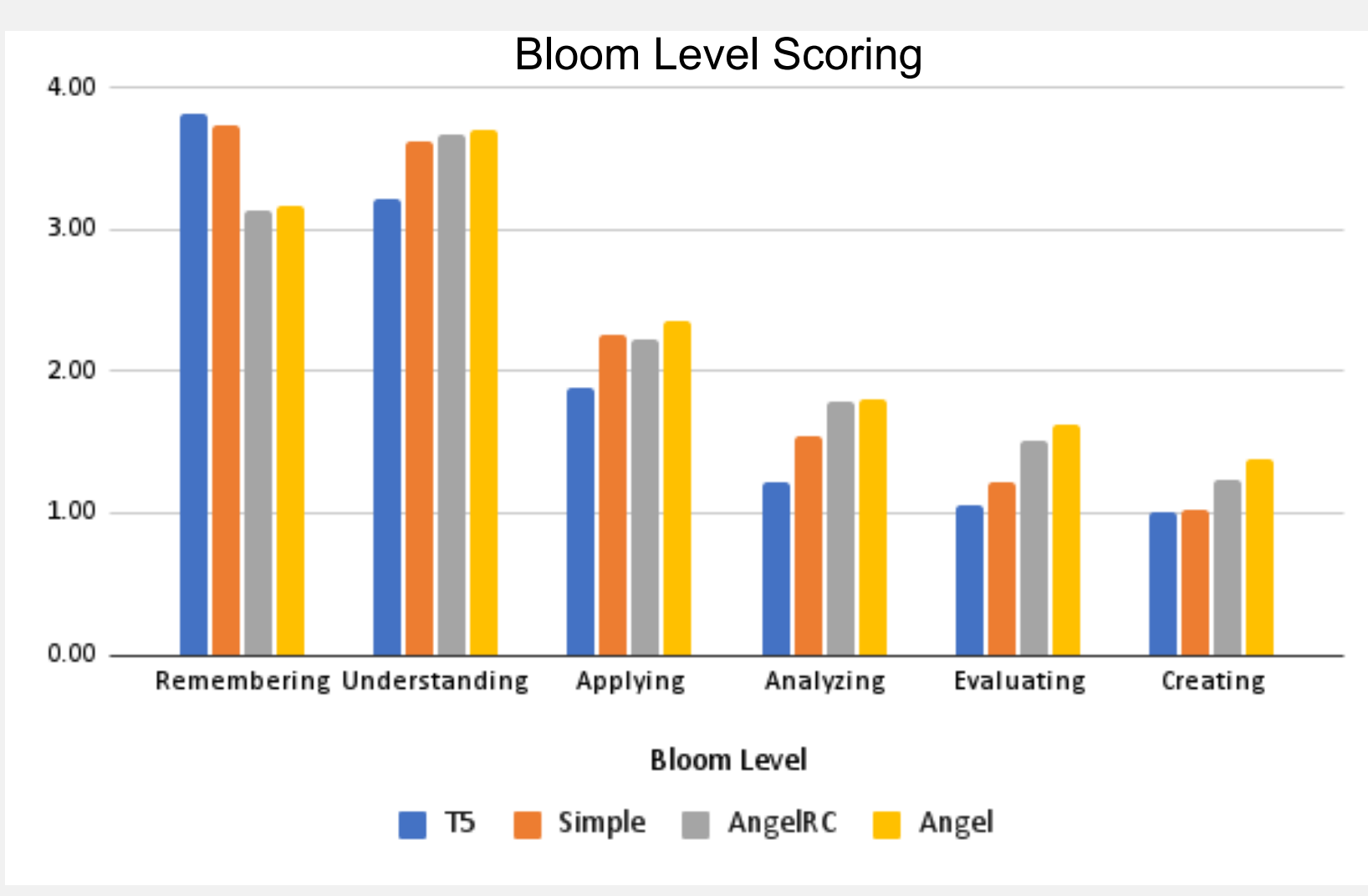
4 Generation Prompt

You are a middle school science teacher. You are given a paragraph from your textbook and you need to write final exam questions and answers for your students. You are supposed to write questions of 3 types: Easy, medium and hard. Additionally, your generated questions should be based on the Bloom Taxonomy. Specifically, you should create more questions with higher Bloom Taxonomy learning objectives (applying, analyzing, evaluating, creating) as opposed to questions with lower learning objectives (reading, understanding). For each generated question, you should specify its main Bloom Taxonomy learning objective and explain why you think this is the question's main Bloom Taxonomy objective. Assume that the following are the Bloom Taxonomy learning objectives: **<omitted>** You can see under "Examples:" below a few examples of past questions created for this paragraph. Please provide each answer with a step-by-step thinking for why the specific answer is the right answer.
Examples: **{Examples}**
The Paragraph: **{Paragraph}**

5 Machine Evaluation

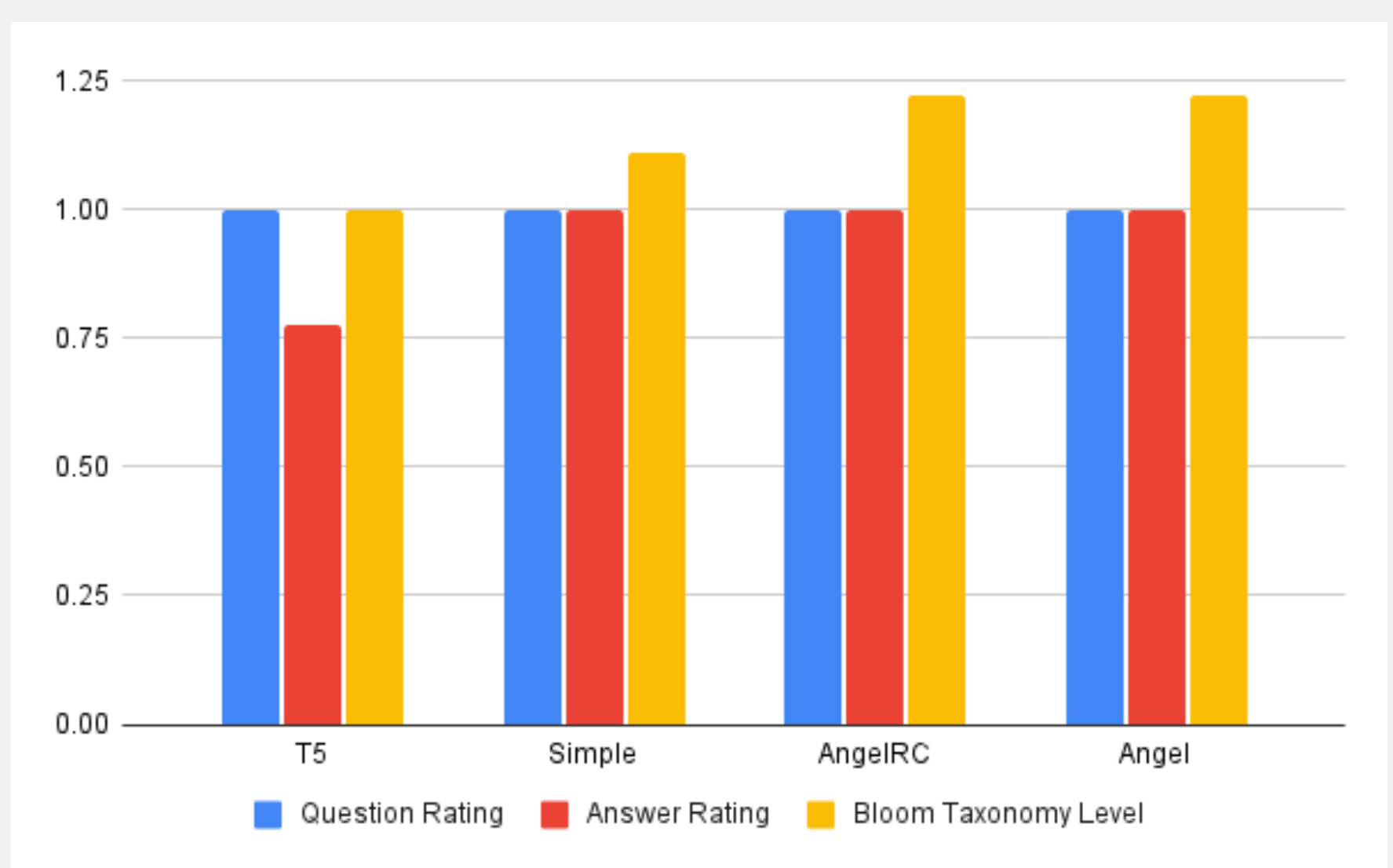
Dataset:

- 8th grade science education book from Indian National Council of Educational Research and Training
- Topics such as crops, coal, plants and animals etc.
- Five to 10 units per chapter
- Generating Q&A per unit



Condition	Clarity	Relatedness	Importance	Answerability
T5	4.63	4.93	4.70	4.52
Simple	4.66	4.96	4.81	4.55
AngelRC	4.51	4.98	4.82	4.29
Angel	4.66	5.00	4.92	4.42

6 Human Evaluation



- Judged Bloom's Taxonomy level
- Rated questions for pedagogical soundness
- Scored answers for correctness

7 Take-home points

- Investigated the use of Gen AI for Q&A generation
- The proposed approach created sound and correct questions and answers
- Integrating educational theory into the generation step showed potential to increase questions' cognitive level
- Future Work:
 - Improve generation and curation processes
 - Test with additional educational sources
 - Increase number of human evaluators
 - Test few shot with higher cognitive level samples
 - Develop multiple LLM based scoring methods