

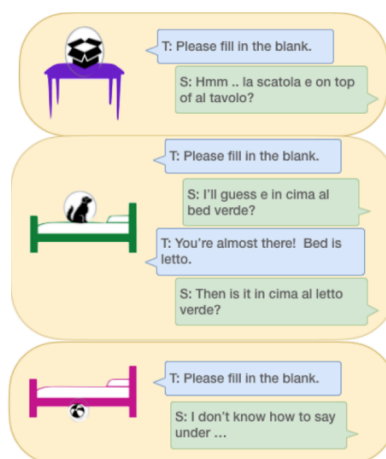
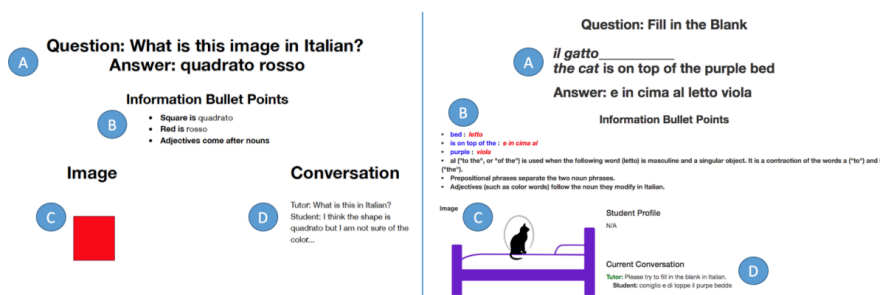
CIMA: A Large Open Access Dialogue Dataset for Tutoring

- Whitepaper: https://people.ischool.berkeley.edu/~hearst/papers/cima_bea_2020.pdf
- GitHub: <https://github.com/kstats/CIMA>

Purpose of CIMA:

Extended conversations are obtained between crowdworkers role-playing as both students and tutors. The CIMA collection, which is made publicly available, is novel in that students are exposed to overlapping grounded concepts between exercises and multiple relevant tutoring responses are collected for the same input. CIMA contains several compelling properties from an educational perspective: student roleplayers complete exercises in fewer turns during the course of the conversation and tutor players adopt strategies that conform with some educational conversational norms, such as providing hints versus asking questions in appropriate contexts. The dataset enables a model to be trained to generate the next tutoring utterance in a conversation, conditioned on a provided action strategy.

Data is made available under the Creative Commons 2.5 license (<https://creativecommons.org/licenses/by/2.5/>).



Student Actions		
Action Label	Description	Example
Guess	The student attempts to answer the question	"Is it 'il gatto e vicino alla scatola rosa'?"
Clarification Question	The student asks a question to the tutor, ranging from directly asking for a translated word to asking why their prior guess was incorrect.	"How would I say 'pink' in Italian?"
Affirmation	When the student affirms something previously said by the tutor.	"Oh, I understand now!"
Other	We allow students to define a category if they do not believe their utterance fits into the pre-defined categories.	"Which I just said."

Tutor Actions		
Action Label	Description	Example
Hint	The tutor provides knowledge to the student via a hint.	"Here's a hint - 'tree' is 'l'albero' because I' ('the') is prepended to the following word when it begins with a vowel."
Open-Ended Question	The tutor asks a question of the student, which can attempt to determine a student's understanding or continue the conversation.	"Are you sure you have all the words in the right order?"
Correction	The tutor corrects a mistake or addresses a misconception a student has.	"Very close. Everything is correct, expect you flipped 'viola' and 'coniglio'."
Confirmation	The tutor confirms a student's answer or understanding is correct.	"Great! Now say the whole sentence, starting with the dog..."
Other	We allow tutors to define a category if they do not believe their response fits into the pre-defined categories.	"Correct! Although try to think of the complete word as 'la scatola.' I find that the easiest way to remember what gender everything is - I just think of the 'the' as part of the noun."

[ASSISTments](#) is a free online math teaching tool developed at Worcester Polytechnic Institute.

- Article: <https://www.govtech.com/education/higher-ed/wpi-researchers-developing-ai-to-provide-homework-feedback>
- Whitepaper: <https://dl.acm.org/doi/pdf/10.1145/3375462.3375523>
- Released Datasets: <https://www.etrialstestbed.org/resources/featured-studies/dataset-papers>

Purpose:

Many systems are either unable to provide support for open-ended problems, relying on the teacher to grade them manually, or avoid such question types entirely. Due to recent advancements in natural language processing methods, the automation of essay grading has made notable strides. However, much of this research has pertained to domains outside of mathematics, where the use of open-ended problems can be used by teachers to assess students' understanding of mathematical concepts beyond what is possible on other types of problems. This research explores the viability and challenges of developing automated graders of open-ended student responses in mathematics.

This data was released by Professor Neil Heffernan of WPI.

N. T. Heffernan, 2014.

Assignment: 6.3 Lesson 5: Comparing Speeds and Prices (6.RP.A.2, 6.RP.A.3.d)

Problem ID: PRABFBR4

[Comment on this problem](#)

Costs of homes can be very different in different parts of the United States.

1. A 450-square-foot apartment in New York City costs \$540,000. What is the price per square foot?

Do not include units (\$) in your answer.

copied for free from openupresources.org

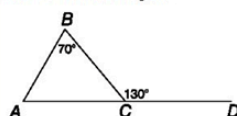
Type your answer below as a number (example: 5, 3.1, 4 1/2, or 3/2):

100% 

Submit Answer

Show answer

What is the measure of angle A?



Original question (mapped to a knowledge component)

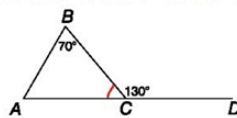
Break this problem into steps

Type your answer below (mathematical expression):

Submit Answer

Sorry, that is incorrect. Let's move on and figure out why!

First you need to find the measure of angle BCA. What do you think it is?



First scaffolding question (also mapped to a skill)

Show me hint 1 of 3

Type your answer below (mathematical expression):

Submit Answer

EdNet:

- Whitepaper: <https://arxiv.org/pdf/1912.03072.pdf>
- GitHub (contains links to datasets): <https://github.com/riid/ednet>

Summary:

EdNet is the dataset of all student-system interactions collected over 2 years by Santa, a multi-platform AI tutoring service with more than 780K users in Korea available through Android, iOS and web.

EdNet dataset contains various features of student actions such as which learning material he has consumed, response, how much time he has spent solving a given question or reading through expert's commentary.

The dataset is publicly released under Creative Commons Attribution-NonCommercial 4.0 International license for research purposes.

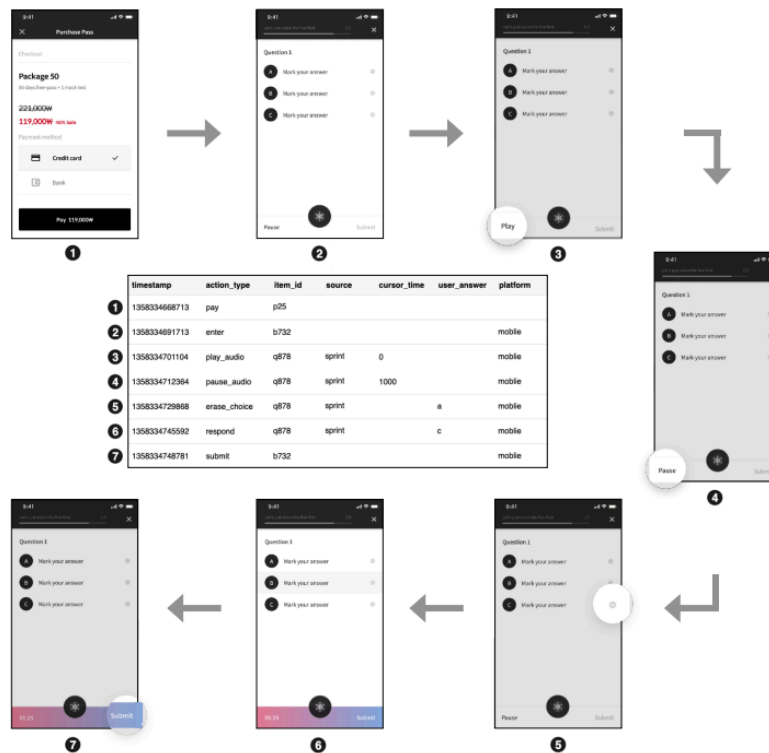


Fig. 1: A possible scenario of a student using *Santa*. After the student purchases a 50-day pass, they solve an LC question. While they solve the question, all of their actions including audio playing and choice elimination are recorded.