Appendix A. Sample Dialogues

The examples of the three dialogue categories were given below.

Gap-clarified dialogue

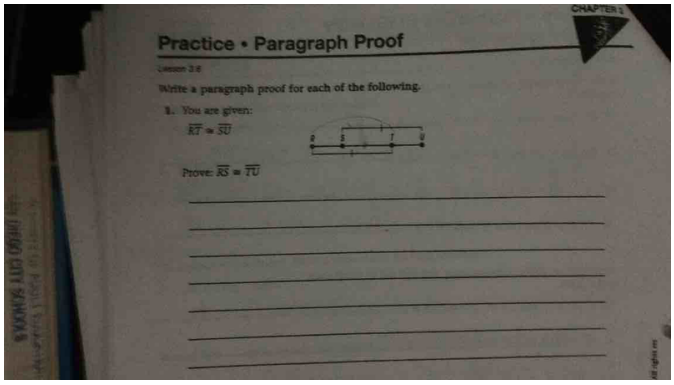

Figure A.1: The Image sent by student in the Gap-clarified session.

Student: [Image Link] (as shown in Figure A.1)
Tutor: Hi [Anonymized]! I'm [Anonymized] and I'll be your tutor for this session.
Tutor: How much progress have you made on that problem?
Student: Well, can you help me with this problem
Tutor: Don't worry:) I will guide you.
Tutor: What would be your best guess as to how we could possibly start the problem?
Student: I have no idea
Tutor: No problem; we will do it together.
Tutor: While solving the problem, we will work together, which will ensure that you are understanding the concept well and in future, you would be able to solve such problems independently.
Student: Ok
Tutor: What information is given here?
Tutor: Any idea?
Tutor: It's mentioned in the question.
Student: Rt ,su
Tutor: Yes
Tutor: RT ≅ SU
Tutor: Well done!
Tutor: Which segment is common to both the parts (segments) RT and SU?
Student: Rs,tu
Tutor: I am asking about the part which is common to RS and TU.
Tutor: Does that make sense?
Tutor: Are you working on the problem?

Gap-explained dialogue

Student: [Image Link] (as shown in Figure A.2)
Tutor: Hello, [Anonymized]!
Tutor: Welcome!

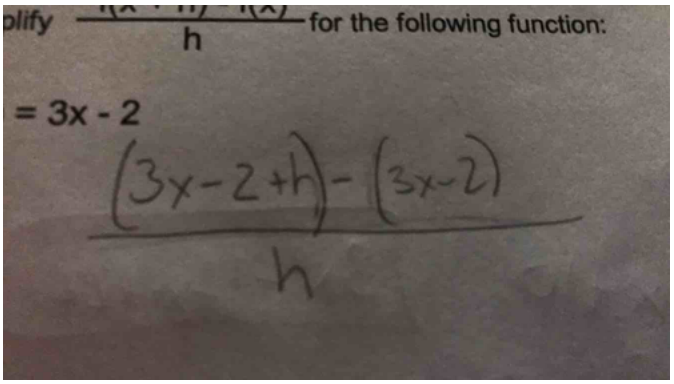

Figure A.2: The Image sent by student in the Gap-explained session

Tutor: This is [Anonymized]! I will be your math tutor for the session.
Tutor: Please give me a moment to go through your question.
Tutor: Given that f(x) = 3x-2 and we need to find (f(x+h)-f(x))/h. Right?
Student: Yes
Tutor: Okay. Thanks for confirming the instructions.
Tutor: How far you have gotten on this problem?
Student: I think I solved it but it has to be wrong
Tutor: No worries! I appreciate your effort.
Tutor: Let's walk through it together and find where is the mistake so that we can correct it.
Tutor: f(x)= 3x-2.
Tutor: So, f(x+h) = ?
Student: (3x-2+h)
Tutor: Not quite right. Let me give you an example.
Tutor: Suppose we have f(x) = 2x+5. To find f(x+3), we replace x with x+3.
Tutor: f(x+3) = 2(x+3) + 5.
Tutor: Okay?
Tutor: Does the example make sense to you?
Student: No
Tutor: Let me show an image.
Tutor: [Image Link] (as shown in Figure A.3)
Tutor: Please go through this image.
Tutor: We have f(x) = 2x+5. When we replace x with x+3 on both sides, we get
Tutor: f(x+3) = 2(x+3) + 5.
Tutor: Then we distribute 2 to get 2x+2(3)+5 = 2x+6+5 = 2x+11.
Tutor: Do you have any doubt?

Gap-bridged dialogue

Student: [Image Link] (as shown in Figure A.4)
Tutor: Hi [Anonymized], welcome!:)
Tutor: Please give me a minute to look over your problem.
Student: Ok thank you
Tutor: How much progress have you made so far? Or are you unsure of how to start?
Student: I know that I have to do 1450g divided by 7.12g but

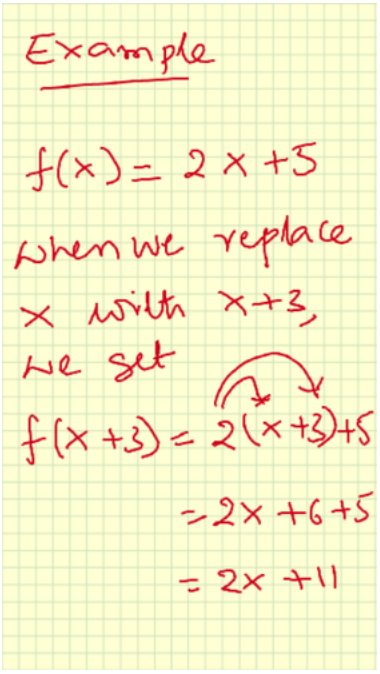

Figure A.3: The Image sent by tutor in the Gap-explained session

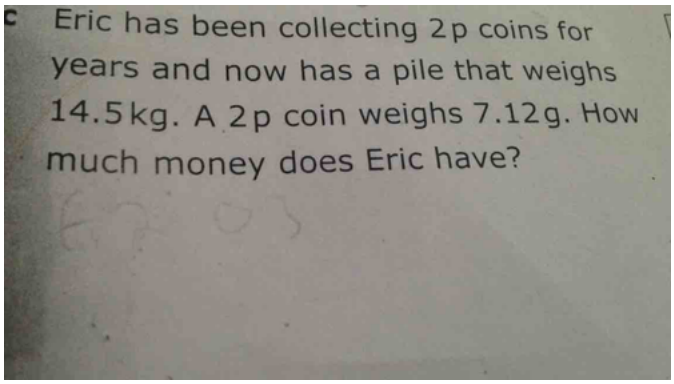

Figure A.4: The Image sent by student in the Gap-bridged session

I'm a bit stuck after that
Tutor: You're nearly there! Looks like you just need some assistance with the unit conversion. I can certainly help you with that. Remember to participate as much as you can and we'll have this in no time.:)
Tutor: Tell me, how many grams are there in a kilogram?
Student: 1000
Tutor: Excellent! So, how many grams would be 14.5kgs?
Student: 14500
Tutor: Exactly! Now, you have got it right. Why do you think we would now divide 14500 by 7.12?
Student: Because one 2p coin weighs 7.12g and we want to find out how many 2p's are 14500g
Tutor: Very good! Alright, so how many 2p coins would be there?
Student: Well 14500 divided by 7.12 is 2036.52 so 2036?
Tutor: Absolutely correct! This is the number of coins that Eric would have.
Tutor: You can also find the equivalent amount of money in pounds, do you have any ideas on how we can do that?
Student: I'm not sure maybe multiplication???
Tutor: Yes, there can be multiplication. How many pounds do you think equal to 1p?
Student: Well there is 100 pennies on a pound so £20.36
Tutor: Awesome! Good job there.:)
Tutor: Thanks for participating. Is there anything else that I can help you with?
Student: Thank you for your help that's all for now but maybe later
Tutor: Sure, thanks for using our service! Have a good one.:)

Appendix B. Ablation Test

To deepen our understanding of the predictive power of each type of features used in the GTB model (when only considering the first 10 utterances as input), we conducted an ablation test, which is a widely used method to evaluate the feature importance on the performance of a prediction model. In an ablation test, the contribution made by a feature is calculated as the difference between the prediction performance of a model when including the feature and that when excluding the feature. The results are provided in Table B.1.

Appendix C. Students Performance Prediction including the student's prior progress

We also experimented by including a student's prior progress in the feature set for problem-solving performance prediction, and the results are shown in Figure C.1.

Table B.1: The ablation test results of the GTB model when only considering the first 10 utterances in a dialogue. Here, Row 1 of of Table B.1 presents the performance of the GTB model taking all available DA and Non-DA features as the input, and Rows 2-13 present the performance of the GTB model without taking a specific type of features as input. The percentages within brackets were calculated by taking the model performance of considering all available DA and Non-DA features (i.e., Row 1) as a comparison. The results with the maximum performance decrease are in bold.

Feature	**Accuracy**	**F1 score**	**Cohen's κ**	**AUC**
1. DA + Non-DA features	0.6805	0.6203	0.3295	0.6470
2. w/o # Individual DA	0.6837 (+0.47%)	0.6231 (+0.45%)	0.3341 (+1.38%)	0.6481 (+0.16%)
3. w/o % Individual DA	0.6801 (-0.05%)	0.6192 (-0.17%)	0.3277 (-0.56%)	0.6462 (-0.13%)
4. w/o # Significant DA patterns	0.6840 (+0.52%)	0.6228 (+0.42%)	0.3354 (+1.79%)	0.6487 (+0.26%)
5. w/o Effort	0.6798 (-0.10%)	0.6201 (-0.03%)	0.3257 (-1.16%)	0.6451 (-0.29%)
6. w/o Informativeness	0.6794 (-0.16%)	0.6199 (-0.06%)	0.3273 (-0.66%)	0.6456 (-0.22%)
7. w/o Complexity	0.6776 (-0.42%)	0.6176 (-0.42%)	0.3227 (-2.07%)	0.6436 (-0.53%)
8. w/o Responsiveness	0.6808 (+0.05%)	0.6221 (+0.30%)	0.3288 (-0.21%)	0.6465 (-0.09%)
9. w/o Questions	0.6815 (+0.16%)	0.6221 (+0.30%)	0.3317 (+0.66%)	0.6476 (+0.09%)
10. w/o Entrainment	0.6794 (-0.16%)	0.6196 (-0.11%)	0.3261 (-1.02%)	0.6451 (-0.30%)
11. w/o Sentiment	0.678 (-0.37%)	0.6184 (-0.30%)	0.3255 (-1.21%)	0.6458 (-0.20%)
12. w/o Experience	0.6837 (+0.47%)	0.6194 (-0.14%)	0.3332 (+1.13%)	0.6478 (+0.13%)
13. w/o N-gram	**0.6762 (-0.63%)**	**0.6109 (-1.50%)**	**0.3126 (-5.14%)**	**0.6384 (-1.33%)**

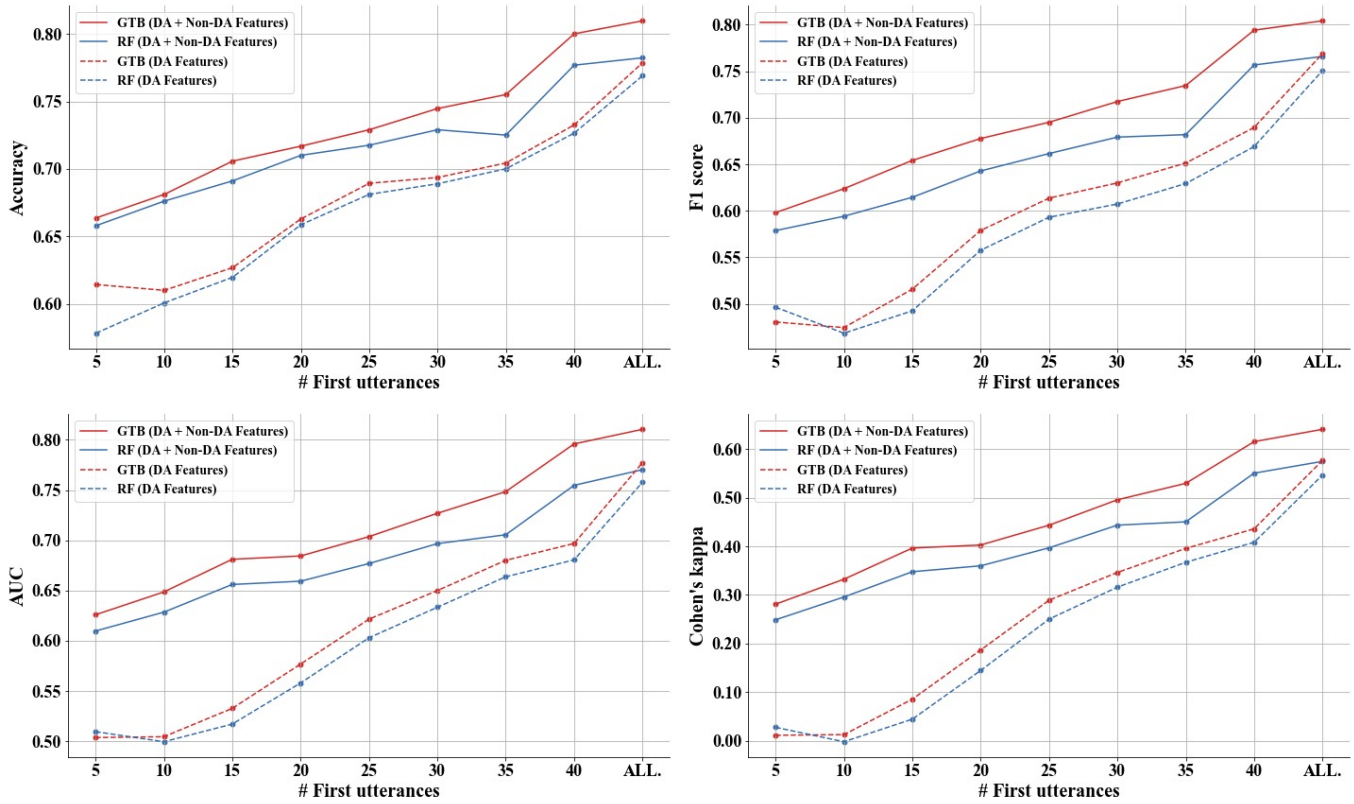

Figure C.1: The performance of GTB and random forests in predicting student performance.