2 ways

1. Score(q) >50 = threshold of a single option (e.g., p > 50). Confusing is when we have at least one option above **50**, otherise non confusing.
/...Score(q) = Sum_over_all_options_above threshold(wi*pi) - Sum_over_all_options below threshold../

~~2. Score(q) = Sum of plausability mass (raw scores) = **200**. Confusing is when we have sum above 200, otherwise non confusing.~~

3. Score(q) = Sum of normalized plausability scores of options that are before the knee point multiplied by the sum of the raw scores of all the options. Threshold is the average knee point of all the knee points questions.

/....Score(q) = Sum_over_all_options_above threshold(wi*pi) ../

4. Totality of explanations: Score(q) = Sum_over_all_options(wi*pi)

- this is weighted sum
- if option is not mentioned in the answer then wi is 0
- wi is a score from 0 to 1

5. Length of explanations

0.15
Normalized Knee Sum x Raw Sum
/ max Raw Sum Value

If Confusing REWARD THIS:
Mentioned Candidates by the test model in the response
/
Distractors
IoU (intersection over Union), Dice coefficient
If not Confusing PENALTY:
1 - (1-0.17)/(1-p_i) for all p_i values =
1 - 0.83/9 =
1 - (1-0.17 + 1-0.15) /(1-p_i) for all p_i values =
1 - (0.83 + 0.85) / 9
1 - 1.68/9 =