

SI 630 Annotation Guideline
Group 7
Assignment 3

We will **label the helpfulness of an answer to a given question on a 5 point Likert scale**: (1) Totally Not Helpful, (2) Not Helpful, (3) Neutral, (4) Helpful, (5) Very Helpful.

Every instance to be labeled will go through three rounds of classification which narrows the label from a range of scores to a single value, followed by a final point adjustment based on the criteria laid out below. The resulting number of points represents the helpfulness of the answer on a 5 point Likert scale.

Criteria 0: Delivery (-ALL pts or nothing)

If the answer is unintelligible, it is considered meaningless and we will deduct all the points.

Deduct all future points and return a helpfulness of 1 if the subject of an answer is unintelligence.

Criteria 1: Subject Agreement

Whether an answer addresses the same realm of topic/subjects proposed by a question is crucial to its helpfulness. If an answer obviously does not have anything to do with the subject of a question, it's content is deemed useless and totally not helpful.

1, 2 = Irrelevant to the subject of the question

3 = Somewhat / indirectly relevant to the subject of the question

4, 5 = Immediately relevant to the subject of the question

The difference between 3 and 4, 5 is that 3 will not give the expected answer for the question. Whenever a question explicitly outlines the format of an expected answer, the style of writing of an answer should conform to the expected format. If the answer doesn't align with the w-words like "what", "why", "when", "where", "how" or something like "ELI5" (which means explaining it in a reasonable way or in plain English), it will be a 3.

If the reply does not address the subject of the question directly, it will result in a 3. However, if the answer is refuting the assumption of the question, it still addresses the

question directly albeit in an argumentative way, thus it still falls in the 4,5 range (eel example).

E.g. [Q]: "Do you think it's weird peoples whole personality is having a big butt?"

- [A]: "I'm not a fan of Instagram for that reason. It's an app designed to be shallow. I like sites and apps that are good for discussion. I don't want to see people bending over or twerking." => **3**
 - (does not address the core question: weird or not weird, but instead gives his opinion which infers agreement in an indirect way)

E.g. [Q]: "Best study habits?"

- [A]: "Kentucky Fried Chicken is my go to restaurant." => **1, 2**
-

E.g. [Q]: "People who prefer soap bars over liquid soap, why is the former your top choice?"

- [A]: "I can eat it if I get hungry" => **1, 2**

E.g. [Q]: "Have you broken the law for a good reason before?"

- [A]: "Nice try FBI" => **3**
 - (inferred that the topic is still on breaking the law)
- [A]: "I have jaywalked to save a kitten before." => **4, 5**

E.g. [Q]: "ELI5: Why do we need to breathe?"

- [A]: "Electron transport chain (ETC)... Anaerobic ... Molecules... (appx 500 words)" => **3**
 - ELI5 implies that the asker expects an answer that uses simpler terms
- [A]: "Because oxygen is required for us to break down particles..." => **4, 5**

E.g. [Q]: "ELI5: How do we still not know how eels reproduce?"

- [A]: "For a long time, it wasn't known how eels mate. Recently, we've observed eels travel to the Sargasso Sea and engage in both courtship rituals and mating. This was only possible thanks to modern GPS trackers." => **5**
 - The answer tries to clarify the fact mentioned in the question and it gives an answer to how we already found out about eels' reproduction process.

E.g. [Q]: "What secret everyone knew except you?"

- [A]: "I dont know they won' t tell me." => **3**
 - This reply is related to the question but does not directly answer the question.

Criteria 2: Organization

The organization of a reply is judged by how well it articulates a claim / response and whether it provides sufficient information to support the claim.

The structure of a reply should follow natural speech patterns where arguments are presented, elaborated upon using evidence, and ultimately elicit meaningful conclusions.

For 1, 2 in Criteria 1:

1 = Little or inefficient use of evidence to support a claim or no structure could be found

2 = Attempts to articulate a claim but may be lacking in structure or evidence that results in confusion

For 3, 4, 5 in Criteria 1:

3 = Little or inefficient use of evidence to support a claim or no structure could be found

4 = Attempts to articulate a claim but may be lacking in structure or evidence that results in confusion

5 = Clearly articulates a claim and demonstrates solid use of supporting evidence

E.g. [Q]: "It's time for the Kentucky Derby - what's the dumbest name you would give your prized horse?"

- [A]: "Gonzo Because, The Kentucky Derby is Decadent and Depraved..." => **5**
 - This is a very tricky one in organization. Firstly, it is obvious that this is very related to the question since a name is given so it will fall into 3 or 4 or 5. Then it will be a 5 because even though this sentence is too short to show his structure, it is very concise and fully expresses what the author wants to say with the story behind.

E.g. [Q]: "What is something you wish you knew when you were in your teens?"

- [A]: "I wish someone had explained community college to me. I had no idea that was a good option until after I was out of school." => **4**
 - This answer is related to the question since a situation is given so it will fall into 4 or 5. There are some illustrations for the reason for this answer. So it will be a 4.

E.g. [Q]: "What was your "I think I left the oven on" instinct or "I got a bad feeling" and was actually true?"

- [A]: "Mostly avoid shitting myself or traffic collisions. The 2 do correlate sometimes" => 3
 - This reply tries to answer the question but it is kind of twisting the question by giving the relation between two stuff he mentioned.
 - It just makes people confused because of two items.

As long as a story attempts to drive a relevant narrative, and covers the key aspects as proposed in the question, then regardless of its length it is considered well organized and will fall into the 5 category.

E.g. [Q]: "To those who've gotten an MRI, what's it like?"

- [A]: "i was told not to swallow, so lots of drool." => 5
 - **Since the answer is trying to narrative the process of this experience instead of just giving the situation**
- [A]: "Lots of drool." => 4
 - **Since the answer is just giving the situation without much narrative.**

Criteria 3: Writing Quality (+0 or -1 pt)

If you need to use Google, check out another link, or do any additional research aside from reading the reply in order to understand any parts that are not disclosed by the question itself, you will deduct a point.

Problematic parts of replies that warrant point deduction: Abbreviation, Proper Noun (People's names, Academic Terminology, etc.)

Whether or not the writing quality of an answer inhibits the reader's understanding of the content also contributes to the helpfulness of said answer.

+0 = Writing does not affect reading comprehension and does not contain parts that needs further research

-1 = Writing negatively impacts reading comprehension due to errors such as misspelled words, grammatical errors, unexplained links, etc. or contains any type of information that needs further research

E.g. [Q]: "What are the secrets to a happy life?"

- [A]: "like people in world, as we can't have fun without no talk to people..." => -1
 - **This is recognizable but it is confusing and needs some guesses.**
- [A]: "Learn to accept that everyone is the main character in their own journey" => +0

E.g. [Q]: "Why would someone like hentai porn?"

- [A]: "[One word](<https://www.youtube.com/watch?v=P2csnVNai-o&t=84s>)" => -1
 - This reply tries to answer the question directly so it is 4 and it will get one point minus because you need to click the link.