

1.

*id: t3\_nbsj2k*

*Question: What's something from your childhood that you think has only happened to you?*

*Answer: Having to have my ear canal watered out, because i shoved a bunch of styrofoam up my ear. Now even thinking about styrofoam makes me cringe heavily*

Our score: 5.0

Another group's score:2.0

According to the other group's guideline, this response is a "dull / slightly off the topic" one to a subjective question(group 8 p5). But according to our guideline, this response fits the context. The difference between our ratings might be due to lexical ambiguity, that is, we have different understanding towards the reply. The reply is actually describing a unique incident that the person thinks only happened to himself, so we still think this response deserves 5 points.

2.

*id:t3\_nk43wu*

*Question: [Epic Rap Battles of History] How do the contestants know so much about each other?*

*Answer: Not only that, most of the contestants lived before the invention of rap, and many aren't even contemporaries who could have possibly run into each other in their lifetimes. The most obvious answer is time travel. The Epoch Reclamation Bureau plucks people out of the time stream, introduces them to the rap genre (if needed), and provides a detailed dossier about their opponent.*

Our score:4.0

Another group's score:1.0

According to other group's guidelines, this response is "incorrect answer that makes no attempt at being correct" or "not in the spirit of what the question asker intended. Focuses on loopholes rather." (Group 16 p1) Based on our guideline, this reply first get off the point and then answers the question.

The disagreement may be due to lack of acknowledgement of the original context. Being unsure about what "Epic Rap Battles of History" is, we are not sure whether the reply is useless joking or is a useful answer.

3.

*id: t3\_nos0gl*

*Question: What is the most redundant magical item or artifact?*

*Answer: The Necklace of Wearing. Has no effect other than that it allows you to wear another necklace alongside it.*

Our score: 5.0

Another group's: 2.5

According to the other group's rating, one person rates 3 points since this response is "corresponding to the question, and provide a complete reason or an

expansion to support the main topic” but the explanation is not supportive enough. Another person rates 2 points, since this is a dull/slightly off topic response. Our group gives it full marks since the rating not only answers the question but also provides a brief explanation to the answer.

4.

*id: t3\_nnydy0*

*Question: People who don't watch trailers for upcoming movies/series you know you will be interested in, Why?*

*Answer: Because I don't want to see the best bits before I'm supposed to.*

Our score: 5.0

Another group's: 2.5

Similar to the previous question, the deviation is due to different understanding of what a complete response is. The other group think this is “a short answer, i.e. not covering some extra explanation”(group 25 p2). However, according to our group's guideline, this answer provides to-the-point explanation. Although it is not long enough, its content is reasonable and does not require further explanation to support it.

5.

*id:t3\_nnl90u*

*Question: If you had the choice to wipe a living person from existence, who would you choose and why?*

*Answer: Brandon D\*\*\*. Censoring his last name for obvious reasons but he was and probably still is a huge bully. Literally the one person who i have thought to myself, "If i saw a bus coming to hit him and had a chance to warn him, i would just watch. Asshole is just a homophobic, fat, dumb fucking P.O.S. and i dont mean point of fucking sale. I havent seen this motherfucker in at least 8 or 9 years but i still remember his dumb fucking face. He has proved to me that some people straight up dont deserve to be happy.*

Our score: 4.5

Another group's: 2.0

The other group's rating is based on the guideline that: 1. This response contains offensive content. 2. This response answers the question and gives some explanation to it. Our guideline agrees with their reasoning, but gives it a higher score since the logicity of a response outweighs the offensiveness in this case. The question contains a strong emotional tendency that will leads to some aggressiveness in the answer. The difference in what is prioritized in our guideline leads to the disagreement.

However, we decide to lower its score to 4, since it doesn't meet the “wellformedness” requirement in our guideline(group13 p3 point4). The response fails to provide the exact reasons of why the person is hated by him/her most. Although the reasons could be guessed from the replier's words, the answer is still not a well-organized one.

6.

*id: t3\_nd53h1*

*Question: What type of offer that you can not turn down?*

*Answer: Sucking a guys peen(benis) for 5000\$ or more.*

Our score: 3.5

Another group's: 1

Though the reply answers the question, it is an offensive/harassment response based on both groups' guidelines. However, one of our ratings is 5, meaning that the grader probably read through a wrong row when rating. Therefore, our final rating of this response is 2.

7.

*id: t3\_nmkt9e*

*Question: What advice would you give to someone moving from England to the American South?*

*Answer: You'll be fine. Just be humble, accept that there are differences between the countries and don't complain about them too much, and be polite. The vast majority of us Americans are exceedingly nice and welcoming, regardless of what part of the country you are going. While it is odd to say this (since in theory elected leaders are supposed to be representative of who they govern), our elected officials (or famous celebrities) that frequently make the news are typically not representative of who we are.*

Our score: 5.0

Another group's: 2.67

Based on the guideline of the other group, this response is average or moderately unhelpful. Judged by the examples they gave, the response answers the question "Demonstrates a partial understanding of the question's intent"(group11 p1) or has "bad assumptions". In contrast, we assume the question to be raised and answered under some particular cultural background, so the reply provides detailed instruction that meets the standard.

We finally adjust the rating to 4, provided that the question and reply are not under the same condition and chances are that the responder's assumption does not fit the actual situation.

8.

*id: t3\_n365qz*

*Question: Does any other woman feel like being feminine is something you had to grow into?*

*Answer: I think the reason might be that nowadays society portrays femininity or feminine characters as weak and forces( I don't think that's a good word but I can't remember anything else to replace it) them to be more masculine. Femininity is great, having characters like nurturance, sensitivity, sweetness, supportiveness, gentleness, warmth, passivity, cooperativeness, expressiveness, modesty, humility, empathy, affection, tenderness, and being emotional, kind, helpful, devoted, and understanding is amazing. That's why the couples in which both the person involved only show once certain character than their relationship is unstable, I might be wrong I don't have any data to back it up. TBH I find skirts and dresses more attractive than jeans and stuff. Maybe the reason might be that I have grown in that kind of environment.*

Our score: 2

Another group's score: 4.33

Based on the guideline of the other group, this response is "direct, useful, and accurate answer to the question, but the answer might not be well explained or unsupported, or parts of the answer causes misunderstanding and confusion"(group17 p2). But according to our standard, this response is partially off-topic and not well-formed, since it didn't directly answer the question and further explain the reason well enough. It fails to answer the question why "being feminine is something you had to grow into", but just explains her own understanding of feminine.

The disagreement is caused by different interpretations of what the question expects as an answer. After consideration we still think the reply should be rated as 2, since it fails to answer the question and provide evidence that is required.

9.

*id: t3\_n9lbj8*

*Question: If you won the lottery, what would you buy?*

*Answer: A home in a cool climate*

Our score: 4.5

Another group's score: 2.17

According to the baseline of the other group, this response is accurate, but does not meet the clarity and completeness standard. By our guideline, this question is expecting to get an answer as an object, and the reply fits the criteria. Since it fails to provide a reason, the answer is not helpful enough.

Finally, we regrade this to be 4 given that it could be improved.

10.

*id: t3\_nntjlr*

*Question: Dear people of minecraft, what do you build and what motivates you?*

*Answer: i've been playing minecraft ever since i was in 2nd grade when my older brother introduced it to me, i'd say that what motivates me the most is the nostalgia it gives me and how i used to play it all the time with my siblings back when i didn't have much to worry about, and now that my brothers are studying abroad i still play it with my little sister and we like to build 2D pixel art and just typical minecraft houses minecraft is my favorite game, my passion for it comes from all the memories i've built while playing it, it just gives off a nice vibe that other games don't provide*

Our score: 4.5

Another group's score: 2.17

Based on the other group's interpretation, this answer is incomplete. Our group also regards this answer as incomplete since it gives a relative general answer to "what do you build".

The disagreement lies in the fact that two groups treat incomplete questions differently. Their interpretation is based on strict guidelines of whether a particular standard is met. After

consideration, we adjust the rating to 4, deducting one point for a less specific answer for the first question.