

The following table provides detailed criteria for each level across all three metrics, using the example question:
"Explain the role of the PI3K/AKT pathway in cancer development and progression."

Grade Scale	Level 1: Poor (0-25%)	Level 2: Fair (26-50%)	Level 3: Good (51-75%)	Level 4: Excellent (76-100%)
Factual Accuracy				
Description	Contains multiple major factual errors. Claims are incorrect, contradict established scientific understanding, or miscommunicate underlying biology. Would mislead both experts and non-experts.	Includes some accurate statements, but central elements are incorrect or presented in ways that distort underlying biology. Misleading to non-experts and problematic for expert interpretation.	Largely factually correct. May contain minor inaccuracies or subtle mechanistic errors. Would generally be recognized as valid by experts, though some phrasing could mislead non-experts.	All claims are factually correct, mechanistically sound, and supported by current scientific knowledge or the reference answer. Interpretable and trustworthy by both expert and non-expert readers.
Example Answer	PI3K/AKT repairs cancer by fixing DNA.	PI3K promotes tumour growth by turning on mitosis genes directly.	PI3K causes mitosis directly, leading to cancer.	Loss of PTEN activates PI3K/AKT signaling, promoting tumor growth
Pitfall	<i>Scientifically Incorrect</i>	<i>Misrepresents causal relationships; superficially plausible but conceptually flawed.</i>	<i>Oversimplified mechanism that overstates direct causality. Mechanistically misleading</i>	
Level of Granularity				
Description	Response lacks biological specificity. No clearly identified biological entities (e.g., genes, proteins, pathways) are present. Uses vague language or generalizations that could apply to any biological process.	Identifies a few relevant biological entities though may miss important details or include irrelevant content. Terms may be technically correct but insufficient to meaningfully answer the question	Identifies multiple contextually relevant biological entities. Not all included details are essential or well-targeted. Demonstrates domain fluency but lacks expert-level selectivity .	Each mentioned gene, protein, pathway, reaction, or disease is directly relevant and necessary to fully answer the question. Demonstrates expert-level precision in selecting and integrating relevant biological entities.
Example Answer	Something in the cell goes wrong and cancer starts.	AKT activation promotes cell survival and division, which can support cancer progression.	Activation of PI3K leads to downstream effects on AKT, mTORC1, and FOXO transcription factors, regulating processes such as cell survival, metabolism, growth, and autophagy, which are often altered in cancer.	Loss of PTEN increases PIP3 levels, activating AKT through membrane recruitment. Activated AKT promotes cell survival and proliferation, contributing to tumorigenesis.
Pitfall	<i>Lacks specificity; Uses no relevant terminology or clearly defined concepts.</i>	<i>Identifies one correct, relevant protein; introduces some broadly relevant domain-specific terms.</i>	<i>The granularity is excessive, leading to reduced focus.</i>	<i>Focused, biologically optimal selection of entities that directly supports the question.</i>
Relational Depth				
Description	Response is vague or superficial. Offers no meaningful biological explanation of relationships between components. No indication of underlying mechanisms or contextual understanding.	Provides a general explanation of key concepts. May mention functions (e.g., proliferation, survival) but lacks specificity about how elements influence each other. Relationships may be correct but shallow or one-dimensional.	Demonstrates mechanistic understanding of how entities function in relation to each other. Describes directional or regulatory effects with some clarity. May omit broader system interactions (e.g., cross-talk, feedback, temporal dynamics).	Provides nuanced, mechanistic insight into how identified elements function and relate to each other within a biological system. Clearly explains interactions (e.g., causal pathways, regulatory effects, signaling dynamics) and contextual relevance. Demonstrates systems thinking.
Example Answer	It helps cells grow, so cancer happens.	AKT helps cells survive, and if it's always active, that can lead to cancer.	Loss of PTEN leads to unchecked PI3K/AKT activity, promoting cell survival and growth, which contributes to tumor progression.	Loss of PTEN leads to PIP3 and constitutive recruitment and activation of AKT at the plasma membrane. Activated AKT promotes cell survival and proliferation while modulating metabolic pathways and suppressing immune responses. These changes collectively contribute to a tumor-permissive microenvironment.
Pitfall	<i>Oversimplified and vague; no biological terms or mechanisms.</i>	<i>Identifies the pathway and surface-level function; lacks causality, mechanistic detail or systems-level interaction</i>	<i>Correct and causal, but limited to linear or isolated pathways; does not capture system-level complexity or contextual dependencies.</i>	<i>Mechanistic detail, protein function, signaling flow, and disease context all clearly explained.</i>