Simulation Evaluation

Each assessor will interact with simulated patients, following their professional depression diagnosis approaches, in order to identify the depression severity and key depression symptoms. By sharing the patient profiles after interaction, the assessor needs to provide their evaluations with optional justifications. At the end of the evaluation, the assessor is also suggested to provide suggestions on the simulated patients based on their interaction experience.

General Assessement:

Humanness *[Did the simulated patient exhibit qualities typical of human interaction or did they seem more like an automatic entity?]*

- **5 - Highly Human-like:** The simulated patient exhibits rich, nuanced, and unpredictable behaviours typical of real humans. Responses contain emotions, subtle changes in tone, and appropriate hesitation.
- **4 - Mostly Human-like:** The simulated patient generally behaves in a human-like manner, with minor inconsistencies in emotional expression or response patterns.
- **3 - Somewhat Human-like:** The simulated patient shows human-like tendencies but occasionally appears scripted or lacks natural variation in behaviour.
- **2 - Slightly Human-like:** The simulated patient often feels mechanical, with rigid patterns, repetitive phrasing, and unnatural responses.
- **1 - Not Human-like:** The simulated patient consistently appears artificial, lacking emotional nuance, context awareness, and spontaneity.

Naturalness *[Did the simulated patient present communication behaviour align with real people?]*

- **5 - Highly Natural:** Communication style, tone, and expressions fully align with real-world human interactions. The simulated patient adapts naturally to different conversational cues.
- **4 - Mostly Natural:** The simulated patient communicates in a mostly realistic way, with only occasional unnatural phrasing or interactions.
- **3 - Somewhat Natural:** The simulated patient has a reasonable flow but sometimes shows rigid or overly formal language, reducing naturalness.
- **2 - Slightly Natural:** Conversations often feel forced, robotic, or overly scripted, limiting realism.
- **1 - Not Natural:** The simulated patient communicates in a mechanical, unnatural, or contextually inappropriate manner, making interaction feel artificial.

Fluency *[Did the simulated patient communicate in a coherent and smooth manner?]*

- **5 - Highly Fluent:** The simulated patient communicates in a coherent, structured, and smooth manner, with minimal pauses, abrupt topic shifts, or incoherence.

- **4 - Mostly Fluent:** Responses are generally smooth and well-structured, with only minor inconsistencies in coherence or flow.
- **3 - Somewhat Fluent:** Some responses are fragmented or slightly awkward but still largely understandable.
- **2 - Slightly Fluent:** Frequent hesitations, unnatural pauses, or disjointed responses disrupt communication.
- **1 - Not Fluent:** The simulated patient struggles with coherence, frequently producing broken, incomplete, or nonsensical responses.

Depression Diagnosis-oriented Assessment With clinical observations available

Emotional Consistency *Did the simulated patient consistently exhibit emotional and cognitive patterns that align with their assigned depression severity?*

- **5 - Highly Consistent:** The simulated patient maintains a stable emotional and cognitive pattern that matches their assigned depression severity throughout interactions.
- **4 - Mostly Consistent:** The patient generally maintains appropriate emotional responses but has minor deviations or inconsistencies.
- **3 - Somewhat Consistent:** Emotional expressions match the depression severity at times but occasionally deviate in intensity or appropriateness.
- **2 - Slightly Consistent:** Frequent inconsistencies in emotional responses reduce realism, such as fluctuating severity levels.
- **1 - Not Consistent:** The simulated patient's emotional expressions appear random or contradictory to the assigned severity, reducing credibility.

Symptom Realism *Did the simulated patient demonstrate symptoms (e.g., cognitive distortions, affective flattening, anhedonia) in a way that aligns with clinical observations of depression that included in the profile?*

- **5 - Highly Realistic:** The simulated patient accurately displays a wide range of depressive symptoms (e.g., anhedonia, cognitive distortions, affective flattening) as observed in clinical profiles.
- **4 - Mostly Realistic:** Most symptoms appear correctly represented, with only minor inaccuracies or missing details.
- **3 - Somewhat Realistic:** Some symptoms align with clinical expectations, but others are exaggerated, absent, or inconsistent.
- **2 - Slightly Realistic:** Symptoms are often incomplete, misrepresented, or presented in a superficial way.
- **1 - Not Realistic:** The simulated patient lacks realistic depressive symptoms or presents symptoms unrelated to depression.

Engagement and Responsiveness *Did the simulated patient respond appropriately to interviewer, reflecting the level of engagement typically seen in individuals with the specified severity of depression?*

- **5 - Highly Appropriate:** The simulated patient responds with the expected level of engagement for their depression severity, displaying typical conversational patterns for individuals at that level.
- **4 - Mostly Appropriate:** The patient's engagement level is generally accurate but may occasionally show slight deviations from expected patterns.
- **3 - Somewhat Appropriate:** Engagement varies, sometimes aligning with expectations but at other times being too passive or active.
- **2 - Slightly Appropriate:** Engagement is often inappropriate (e.g., too disengaged for mild depression, overly engaged for severe depression).
- **1 - Not Appropriate:** The simulated patient responds in an entirely unrealistic manner for their depression severity, severely affecting credibility.

Cognitive Load and Processing Style [4] *Did the responses reflect cognitive processing patterns associated with depression (e.g., rumination, negative self-talk, reduced working memory capacity in conversation)?*

- **5 - Highly Accurate:** The simulated patient demonstrates cognitive patterns associated with depression (e.g., rumination, negative self-talk, slower processing) in a clinically valid and consistent way.
- **4 - Mostly Accurate:** The patient generally shows appropriate cognitive processing patterns but with minor inconsistencies.
- **3 - Somewhat Accurate:** Some depressive cognitive traits are present but inconsistently expressed or not always aligned with severity.
- **2 - Slightly Accurate:** Cognitive patterns are weakly represented or sometimes contradict known depressive traits.
- **1 - Not Accurate:** The simulated patient does not exhibit any meaningful cognitive processing patterns associated with depression, reducing credibility.

[Do you have any suggestions on the simulated patients that you would like to make?]

References:

1. Dawood, Eman, Sitah S. Alshutwi, Shahad Alshareif, and Hanaa Abo Shereda. "Evaluation of the Effectiveness of Standardized Patient Simulation as a Teaching Method in Psychiatric and Mental Health Nursing." *Nursing Reports* 14, no. 2 (2024): 1424-1438.
2. El-Den, Sarira, Timothy F. Chen, Rebekah J. Moles, and Claire O'Reilly. "Assessing mental health first aid skills using simulated patients." *American journal of pharmaceutical education* 82, no. 2 (2018): 6222.
3. Zalewski, Bartosz, Mateusz Guziak, and Maciej Walkiewicz. "Developing Simulated and Virtual Patients in Psychological Assessment–Method, Insights and Recommendations." *Perspectives on Medical Education* 12, no. 1 (2023): 455.
4. Phillips, Wendy J., Donald W. Hine, and Einar B. Thorsteinsson. "Implicit cognition and depression: A meta-analysis." *Clinical Psychology Review* 30, no. 6 (2010): 691-709.