Supplemental Materials

Results

System Usability Scale

Positive SUS questions assess aspects such as **ease of use**, **functionality integration**, and **user confidence**. Observations indicate that questions like "The system was easy to use" and "The functions in this system were well integrated" received moderate scores, reflecting general satisfaction with these aspects. However, lower scores for "I felt very confident using the system" suggest that while users found the robot crash cart functional, they lacked assurance during use. Additionally, relatively low scores for "I think that I would like to use this system frequently" indicate limited perceived usefulness or appeal for repeated use.

In contrast, the negative SUS questions highlight areas where the robot crash cart **struggles in usability**. High scores for "I found the system unnecessarily complex" and "I needed to learn a lot of things before I could get going" reflect significant challenges in robot crash cart intuitiveness and initial onboarding, while higher scores for "I think that I would need the support of a technical person to use this system" suggest that users may require external assistance to effectively navigate the robot crash cart. Moreover, moderate scores for "I found the system very cumbersome to use" indicate usability issues that detract from user efficiency and satisfaction when using the robot crash cart.

The positive SUS questions (see Figure 1), which focus on usability of the robot crash cart, show a range of average scores between 2 and 3. The lowest average score was 2.0 for "I felt very confident using the system" (Q9). On the other hand, the highest average score was 3.0 for "I would imagine that most people would learn to use this system very quickly" (Q7), suggesting that users believe the robot crash cart's learning curve is manageable for most people. The standard deviations (STD) vary, with Q1 ("I think that I would like to use this system frequently") showing the highest STD at 1.33, while Q9 ("I felt very confident using the system") has the lowest STD at 0.87. Overall, the positive scores reflect moderate satisfaction but highlight a lack of confidence in using the robot crash cart.

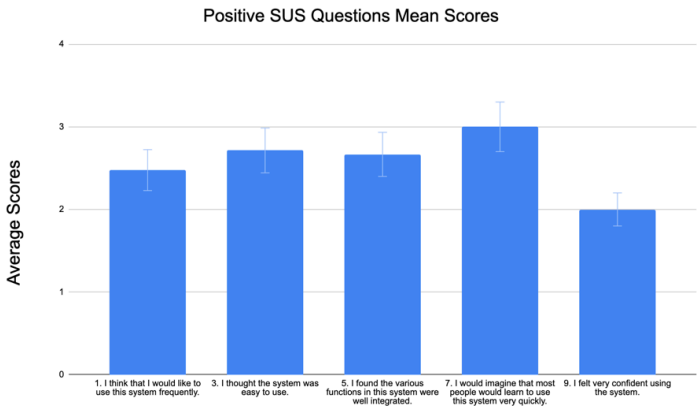

Figure 1: Positive SUS Questions Mean Scores

In comparison, the negative scores (see Figure 2) were generally higher than the positive scores, suggesting more dissatisfaction than satisfaction in usability among users. The negative SUS questions (see Figure 2), which assess areas of complexity, inconsistency, and difficulty, resulted in averages ranging from 2.76 to 3.71. The highest average score was 3.71 for "I needed to learn a lot of things before I could get going with this system" (Q10), indicating a strong perception that the robot crash cart requires significant improvement in terms of user training before interacting with the robot. In contrast, the lowest average score was 2.76 for "I found the system unnecessarily complex" (Q2), suggesting that while the robot crash cart is perceived as complex, it may not be the most pressing issue. These questions resulted in the highest and lowest STDs which are relatively similar, ranging from 0.93 to 1.15. Overall, the negative scores highlight difficulties with usability and training users to use the robot crash cart as a learning tool. The gap between the highest negative score (3.71) and the lowest positive score (2.0) emphasizes that users struggled significantly with the robot crash cart preparation and confidence which are significant usability concerns.

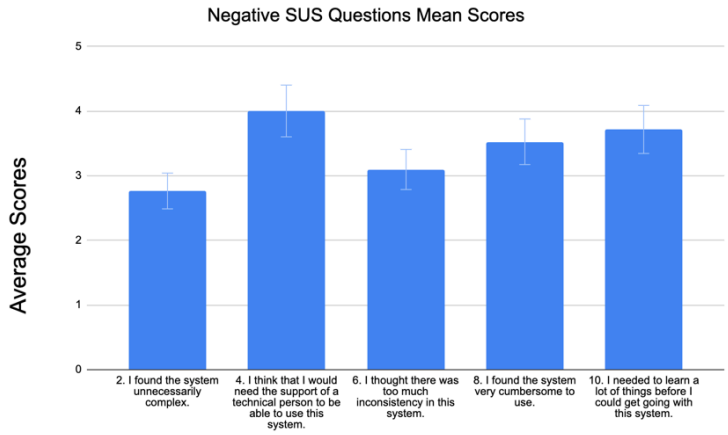

Figure 2: Negative SUS Questions Mean Scores