Over the past few days, Dr. Man and I've made strong progress on our eye-tracking and machine learning integration project for spatial reasoning testing. First, we completed **data preprocessing**, including handling missing values, normalizing skewed distributions using log and power transformations, and standardizing features for modeling. We explored all eye tracking metrics and find key metrics like fixation duration, pupil size, and saccade amplitude.

Next, we conducted **unsupervised learning**, using PCA, t-SNE, and UMAP for dimensionality reduction, followed by KMeans clustering. Although visual cluster separation was limited, this helped us understand that performance differences might not be cleanly separable through gaze data alone.

On the **supervised side**, we trained Random Forest models using binary accuracy labels, IRT-based latent ability, and cognitive strategy label. We identified several important predictors—particularly fixation count, saccade count, pupil size—as strong signals of performance.

These empirical steps fed directly into refining our **research questions**, which now focus on how gaze behavior distinguishes piecemeal and holistic strategy participants, and also distinguishes high- and low-performing participants, how real-time attentional states could personalize item delivery, and whether incorporating behavioral data helps mitigate guessing and disengagement in spatial test.