Implementation Ideas and Assumptions:

1. Participant will basically sit straight during the test. Participant will not move his/her body too much during the test. Participant may move slightly with head and body during the test and it is tolerable. The program constantly detects face and update distance between nose and camera.
2. The camera will be set up in front of the participant at height lower than the participant’s body. This makes the camera looks up at participant. When participant put index finger in front of face, the finger tip can be detected in depth input.
3. It’s very hard to distinguish finger from face using color input alone when finger is put in front of face. Because both parts are human fresh color. However, by using depth input it’s easier to separate finger from face as the two parts have different distance readings to camera.
4. OpenCV face detection can work well in low illumination situation. And depth input won’t be influence by environmental illumination or skip color or cloth color.
5. The prototype uses color input to detect the face of the participant. (Currently using OpenCV cascade.) Then it considers the middle point of the face as nose. Then uses x and y information of nose to get the distance between nose and camera from depth input. This distance will be used as depth input cut off threshold. The program assumes participant puts hand and finger in front of face.
6. The calculation of nose distance is performed by at most 9 nearby points around the nose and take the average.
7. The lower body of participant could be inside the depth cut off threshold. (Body could be seen in front of face.) Hence, the program uses a calibration phase to detect body part and block that part (denoted by white rectangle in the prototype).
8. In order to reduce CPU usage for face detection. Only central part of the color input is used to detect face (denoted by green rectangle in the prototype).
9. Image resolution and frame rate can be adjusted using global variables in the prototype. For Intel RealSense D435 Camera, it supports 848 x 480 @ 60 fps and higher resolution with 30 fps.
10. The primary processing of depth input includes threshold cut off (remove everything behind face), contour detection, medium blur, small area removing (the real hand/finger part should have a relative considerable size), and convex hall.

