CS\_Moreland Application Comments by IMMD.

The main comments are (in ethics application form):

1. project description: the justification of collecting personal recordings is to perform comparison and provide suggestions to users, but this has not been discussed clearly in Q31: how the comparison and suggestion will be performed?

>> *Comparison of anonymised (true) recordings and user/tester recordings will be done through Google speech-to-text API initially, then normalised frequency distribution and histogram analysis for feedback. Advanced comparisons could use phoneme detection but this is unlikely in the timescales allowed.*

*The goal is to have the user able to voice the complex pharmaceutical name in a “similar” or “correct” pattern – and these terms have to be defined for measurement.*

*Now covered in Q31 under the heading “Project Description”*

2. Q30 a: what is the secure communication channels? how to ensure the data not be shared outside professional use? Here we look for concrete mechanisms to address these concerns, not promise.

*>> medical practitioner data through transferred through WhatsApp’s 5 layer encryption or university’s OneDrive and stored on student’s server with access through the app which is not shared widely but only with test users. Sharing on a technical level will be either contained on a local device without transfer or will be with Google and the participants are informed of this.*

*Given the timescale left we cannot have a full trial with medical students so I will have to address this issue in a later project – which will incorporate access to a shared university resource with recordings either encrypted or under access only with a university id.*

*For Joanna’s project I can only see that she and I will test the efficacy of matching histograms as proof of concept. The data will only be accessible by Joanna and myself.*

*Discussed in Q30 a, b and c*

Q30 a: what do you mean by “users will have access to anonymous recordings”? In the first comment, we have mentioned that recordings are person identifiable,  and how do we protect person’s identity? Why the other users even need to have access to others recordings?

*>>This refers to the practitioners voice recordings – Joanna does not know the people concerned and once we gather the recordings they will be stored under nominal descriptors such as <male, native speaker> or <female, non-native speaker>.*

*Participants sign off on agreement to use their voice recordings and identifiability. Labelling will not identify practitioner.*

*User accesses voice recordings to imitate in order to learn pronunciation.*

*Discussed in Q31 “Project Description”, “Informed Consent”, “Risk of Loss of Confidentiality of Participant Information” and “Address Risks”*

Q30 b: when you mention “partipicants’ data”, you might need to be very specific: because you collect recordings and questionnaires, what the “participants’ data” is? Do you refer to both?

*>> Q30 b has now clarified whose data*

In the application, we look for the clarification of the following points:

1) what data do you collect?

*>>practitioners send recordings through WhatsApp. Practitioners submit voice recordings. Users/testers submit voice recordings. Their voice recordings are not stored past the comparison with practitioner’s recording. Survey results. All data (practitioner recordings and survey results) destroyed post project.*

*Covered in Q30 a*

2) why do you need to collect these data?

*>>For comparison against the professional recordings for training purposes and for user/tester feedback. With the aim of allowing testers to practise their pronunciation of complex pharmaceutical and medical terminology against practitioners.*

*Discussed in Q31 “Project Description”*

3) is there any ethical risk to the data, to the participants? If so, what they are.

*>> Addressed in Q31 “Risks” and “Risk of loss of confidentiality of participant information”*

4) how do you address the risk?

*>> Solution is to store on a secure server and delete after project completion.*

*Recordings stored on university server. Survey data also stored and anonymised. For project length also stored on student laptop which is password protected.*

*Discussed in Q30 c and d and Q31 “Address Risks”*