# Piloting with the PSUM Scale: Establishing Usability First

**S. Greenstein, B. D. Sawyer, E. Niederman, P. Oppold, and P. A. Hancock**

University of Central Florida, Orlando, FL, USA

Institute for Simulation and Training

Department of Psychology

## Introduction

- Rapid development of vehicle-to-vehicle (V2V) tools creates the challenge of determining usability (Lavrinc, 2012)
- Population Specific User Mastery (PSUM) Scale as research measure for driving research (Sawyer et al., *in press*)
  - Usability of a tool by a sub-population may connote a research confound
  - Three cell phones were evaluated in the piloting before a study
- PSUM Scales help observe differences in a population using a chosen tool
  - Where do training or design changes need to be implemented?
- A PSUM requires minimal information
  - A population, a tool, a series of tasks within that tool, and a time criterion
- A PSUM Score is the percentage of the measured population able to accomplish the evaluated task within the criterion time

## Method

### Materials

- iPhone 4S
- LG P509 running Android 2.5
- Motorola XT907 running Android 3.0
- Stopwatch

### Participants

- 16 undergraduate students from the University of Central Florida
  - Ages 18-20
  - Data from 4 participants was not used due to technical difficulties.

### Procedure

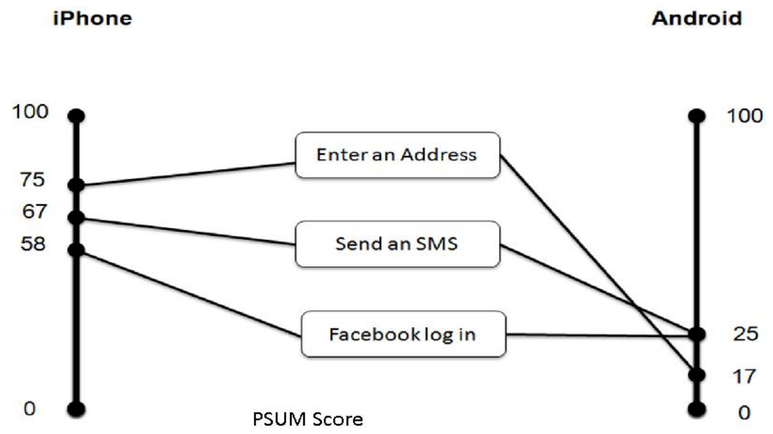
- Time taken by participants to complete 10 different tasks was recorded
  - All tasks were completed on both devices
  - Tasks included: turning the device on and off, sending an SMS message and using the device's navigation system
- Demographics questionnaire

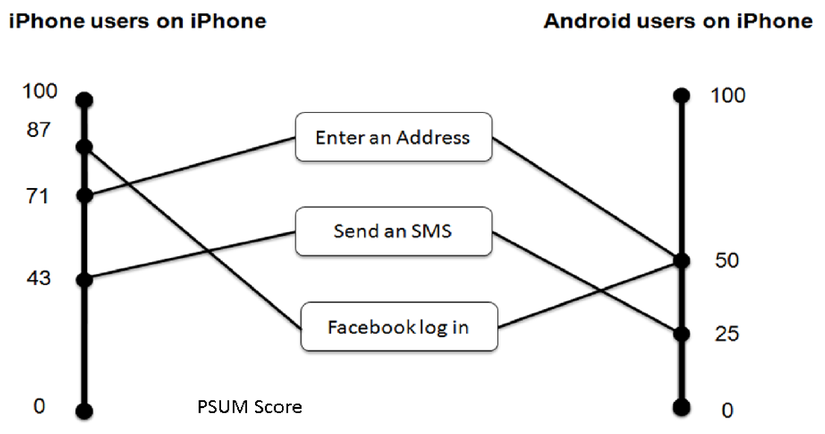
### Research Questions

- iPhone or Android?
- Will participants need training?

## Results

- More tasks were completed within the criterion time on the iPhone than the Android
  - Navigation for Android ($t(11) = -2.05$, $p = 0.05$; $M = 83.81s$; $SD = 85.11s$); iPhone ($M = 28.50s$; $SD = 10.46s$)
  - SMS messaging on the Android ($t(11) = -2.11$, $p = 0.05$; $M = 48.01s$; $SD = 21.64s$); iPhone ($M = 32.46s$; $SD = 10.82s$)
  - Facebook application on the Android ($t(11) = -2.66$, $p = 0.02$; $M = 67.87s$; $SD = 47.13s$) ; iPhone ($M = 35.28s$; $SD = 15.74s$)
  - The two Androids showed negligible differences in task completion, except for turning off the device.
    - LG P509 ($M = 16.37$; $SD = 5.09$) and Droid Razor M ($M = 40.19s$; $SD = 29.94s$)
- iPhone users on the iPhone still had some trouble completing our tasks within the criterion time
  - Android users on iPhone were substantially worse off
- As a result, a 15 minute training session for all participants was implemented in the final study.




## Discussion

- When comparing in-vehicle devices no standard metric exists
  - To compare performance between populations
  - To compare tools within a population
- The PSUM Scale provides such an objective metric.
  - Using information already available in most pilot studies, a constructed PSUM scale guided two decisions about the final study. **First**, the decision was made to use the iPhone. **Second**, 15 minutes of training was provided to each participant.