Introduction

The following document presents the External Audience Protocol for testing and evaluating smartwatches at Consumer Reports. The purpose of this test process is to expose the performance strengths and weaknesses for smartwatches, and to make a final assessment of its overall performance. A high level description of our test process can be summarized by the following attributes:

1. Step-count accuracy
2. Heart-rate accuracy
3. Ease of Use
4. Water resistance
5. Display scratch resistance
6. Versatility

Description of Test Process

Each category's test process is a tabulation of subscores that include, but are not limited to, the examples below:

1. Step-count accuracy

   This attribute grades the accuracy of steps counted compared to a reference using a group of trained testers.

   A group of trained testers composed of two (2) men and two (2) women of various heights, weights, and ages walk two (2) runs on the same route. Each sample is paired with a recent version smartphone with the manufacturer’s recommended default app installed. For each sample, individual statistics (height, weight, gender, age) are entered into the app’s profile for each tester.

   Each tester walks the test route. At the end of each run, the measured step count data from the smartwatch is read at-wrist along with the data from the manually operated Hand Tally reference. If the raw data is consistently off by ten (10) or more steps when compared to the manual counter, it is retested to be sure that it is not a mistake and to see if optimized results can be obtained.

   Each smartwatch is tested in accord with the manufacturer's instructions for measuring steps. However, it is possible for the tester to misuse the product by wearing it too loosely, which could cause inaccurate readings. It is most accurate
when the arm motion is continuous, repeated in the same direction, and at the same pace. So, strict attention is paid to the location and tightness of fit on the wrist. Smartwatches are worn according to the manufacturer's instructions on the non-dominant wrist and the manual counter is operated using the dominant hand. The training is specific about the route, so there is no variation in the path each tester takes while walking the test route.

The statistics department analyzes the raw data to calculate LSMeans and then ratings scores are produced.

2. Heart-rate accuracy

This attribute grades the accuracy of heart-rate measured during exercise at various stress levels as compared to a reference chest-strap sensor worn by each of a group of trained testers.

The group of trained testers is composed of two (2) men and two (2) women of various heights, weights, and ages. Each doing two (2) runs on a stationary bicycle. Each sample is paired with a recent version smartphone with the manufacturer’s recommended default app installed. For each sample, individual statistics (height, weight, gender, age) are entered into the app’s profile for each tester.

A stationary bicycle is used for testing. Tests will occur at four (4) speed conditions: Rest (<= 75 bpm): Walk (110 bpm), Jog (130 bpm), and Run (140 bpm). A run consists of recording a data point at all four (4) speeds, so there are four data points per run. There are two runs per tester so eight (8) data points are recorded for each smartwatch along with the eight (8) data points from the reference for comparison.

The raw Heart-rate data is submitted to the statistics department for analysis and they create a report based on their statistical design. The raw data is then compared to the reference and analyzed by the statistician to produce a Heart-rate Accuracy Statistical Report with LSMeans that are used to create ratings.

If the heart-rate measurement data is consistently outside the expected limit of greater than or equal to fifteen (15) bpm difference from the reference or if there is an error, or no explanation for unusual results, a re-test is performed. The Statistics Department, under certain circumstances may require a retest of the same sample or a retest using a second sample.
Where possible misuse is concerned, it is possible for the consumer to misuse the product by wearing it too loosely which could cause the sensor to take inaccurate readings due to insufficient contact between the Smartwatches optical heart-rate sensor and the tester's skin. The measurements are most accurate when the optical heart rate sensor maintains secure contact with the tester’s skin. This is mitigated as much as possible by paying strict attention to the location and tightness of fit on the wrist. Each smartwatch is worn on the non-dominant wrist to maintain contact with the sensor throughout testing according to the manufacturer’s recommendations.

3. Ease of use

This attribute grades the ease of using the button(s), rotating crown/bezel, menu navigation controls, touchscreen responsiveness, viewing the watch face, ease of pairing, ease of interaction, how easy it is to set up a new device using the default apps, and more.

All of the steps required for setting up and operating the models are evaluated as to how easy or difficult they are to use by two trained experienced testers. This, coupled with a combination of questions, pedigree items, and scenario based evaluations will provide the EOU score for ratings. Questions are based on discussions at meetings regarding the most commonly used features with a focus on consumer interaction. Meetings were held with department staffers from Electronics, Consumer Engagement, Editorial, Statistics, and Management to ensure that EOU will be consumer centric in its approach. A usability study coupled with user surveys were also used to inform.

The overall Ease of Use score is made up out of the following weighted attribute EOU subscores:

- Ease of interaction
- Ease of pairing
- Bright Light Text
- Bright Light Time
- Low-Light Text
External Audience Protocol (EAP) - Smartwatches

- Low-Light Time
- Ease of connecting to charger
- Charging connector detaches unintentionally or is much less secure than most others
- Ease of charging
- Rotating Bezel navigation
- Rotating Crown navigation
- Button based only Wrist-raise turns screen on
- Always-on display capability
- User interface has "intuitive functionality"
- Has Standalone GPS
- Initial learning curve - trial and error
- Has insufficient printed user manual instructions
- Has insufficient Online user manual instructions

4. Water resistance

Smartwatches are tested in a pressure chamber according to each manufacturer’s claimed depth and duration. A two-out-of-three test methodology is employed. So, a sample has to fail twice to be designated a failure. If there is no claim, each smartwatch is subjected to a baseline test. A passing or a failing score is generated.

5. Display scratch resistance

This attribute grades the ability of the watch face to resist being scratched using mohs hardness tips and the mohs hardness scale results are used to develop a score.

6. Versatility
This attribute grades the presence of typical useful features, such as: multisport tracking, move bar, OS compatibility, find my watch, wrist-raise turns screen on, floors counting, NFC, cellular, touchscreen, standalone GPS, socialization, notifications, and more. A final score is generated.