1. The droids are captured and not there voluntarily, and perhaps Vader's delivery is worse. You'd need to have more willing participants, and use the same representative or both together.

2. The samples were not similar (unhappy vs. happy). The representatives could have done different jobs as well at the deliverance. You'd have to pick a random sample from each country and do them separately, also only use one representative or together tackling a group at the same time in the same country.

3. The countries are all different and the distributions are scattered, not enough similarities between samples. You could take samples from each country and compare them to each other, not each country against the other.

4. There's no baseline to tell you how active they were before they downloaded the app. They could have all users take a survey as to how active they were before the app.

5. The teacher did not distribute the tests A,B,C. Kids with test A were next to each other, B, C. She should distribute the tests randomly.