Challenge: Evaluate an experiment analysis

1) This experiment is designed in a way that could produce bias in several ways. There are power dynamics that can influence the response of the droids. First, the droids are captured which implies that they are unwilling participants in the study and that they are subject to influence, resulting in a bias response. It is not a double blind study. Therefore The Emperor and Darth Vader are aware of how each droid responds to the question. Considering their positions of power, vis-à-vis the droids responses by the droids my not reflect their true sentiment. The sampling procedure does not appear to be random. If this a recruiting slogan, most likely it is designed to persuade entities other than the captured droids? Who is their primary audience? The sample could be significantly different from the populations. Additionally we don’t know the demographics of the captured droids. Are some of the droids from a place that is more likely to support one slogan over another?

To improve this experimental design they should make sure that the sampling procedure is random and representative of the population. Captured droids are not a good sample population. The experiment should draw from a sample population that may not risk repercussions. They should ensure that if there are sub-populations in the group who may have a bias for or against a certain slogan, that that sub-population does not bias the results. They should develop a double blind experiment so that Darth Vader and the Emperor are not involved in the experiment and do not know who is converting.

2) With this study design it is difficult to differentiate between the sample population and the population – and its randomness. The groups may not be as similar as is assumed in the study. The study context can create bias. What constitutes a ‘friendly’ planet and an ‘unfriendly’ planet? Do all the citizens of the friendly plant agree with the Jedi? Or do some citizens on these planets hold friendly *and* unfriendly feelings on each planet. Perhaps, 25% of the population on one ‘unfriendly’ planet is friendly to the Jedi and 45% of the population of another ‘unfriendly’ planet is friendly to the Jedi. If the Jedi has difficulty with public relations then one can deduce that there are unfriendly citizens on all planets. We also don’t know the population size of these planets. Are there sample sizes quantitatively different? If so this will influence the results. In this design, there is a risk of Simpson’s Paradox. There may be several lurking variables that would complicate the conclusions of the study.

To improve this experiment the sample population should be carefully constructed. They should make sure that plants are similar enough so that one planet that may have demographic or other dynamics that differ significantly from the others does not lead to a skewed result. They should also make sure that outside factors do not influence the results of their survey. Were there other factors such as ads, literature etc. that have led to some bias in the response of the survey respondents?

3) The two groups are not comparable. The study refers to job type by location. The location can have a strong influence on job satisfaction rates. The location comprises many factors, including hierarchy in the workplace. There are many lurking variables that this design does not capture and could easily fall victim to Simpson’s Paradox. In this design they are only asking about HR and IT workers without differentiating the experience of workers within each group. For example, gender could influence job satisfaction; job ranking could also influence satisfaction. How might salary differentials influence their job satisfaction?

To improve on this design the authors of the study should consider developing several metrics to measure satisfaction to capture how patterns amongst sub-groups may produce results that show opposite and opposing trends.

4) This experiment should be designed to capture behavior over time. The data show an increase in activity does not take into considerations spikes and declines of activity that could happen overtime. The novelty of the app rather than the app itself could be causing an increase of activity. Without a long term analysis the effectiveness of the app could be over estimated.

To design a better experiment the company should adjust its research question. Rather than analyzing the immediate response to the app, they should collect data for at least one year to see if over time activity levels continue to rise. Or if the spike in activity is a phenomena linked to the sampling procedure. There should also be an analysis of the people who choose to ‘opt in’. Are they people who were downloading the app because they want to start a new fitness regimen, they are making a New Year’s Resolution, etc? A short survey asking participants why they chose to download the app would be useful for the study.

5) The teacher designed a random survey of the students by staking the tests and randomly giving the test to students. But the teacher cannot conclude that test B is easier. There could be lurking variables that could explain why students taking test B scored better. On average the students taking test B scored higher but is there an (or several) outliers that influenced the average score? Did students taking B have a greater number of high scores than students taking A or C? If these outliers were removed would the average of be more similar to A and C?

To improve on the design of the experiment, it would be in the teacher’s interest to repeat the study with other classes several more times to get more accurate results. The teacher should pay particular attention to the median and standard deviation of the test results to have a better understanding of how the design of the test may be related to the test results.