Now it's time to flex your critical evaluation skills. Read the following descriptions of an experiment and its analysis, identify the flaws in each, and describe what you would do to correct them.

1. The Sith Lords are concerned that their recruiting slogan, "Give In to Your Anger," isn't very effective. Darth Vader develops an alternative slogan, "Together We Can Rule the Galaxy." They compare the slogans on two groups of 50 captured droids each. In one group, Emperor Palpatine delivers the "Anger" slogan. In the other, Darth Vader presents the "Together" slogan. 20 droids convert to the Dark Side after hearing Palpatine's slogan, while only 5 droids convert after hearing Vader's. The Sith's data scientist concludes that "Anger" is a more effective slogan and should continue to be used.

* **Flaws**
  + Droids aren’t the greatest representation of the population as the star wars universe has many different sentient species. Droids are the worst pick as they are synthetic and the least representative of the population of the galaxy.
  + A sample size of 50 may not be enough to represent the diversity in a galactic-scale population.
  + There is bias introduced because one group is addressed by Vader while the other is addressed by Palpatine.
  + Both Vader and Palpatine are bad choices as they are pretty intimidating figures and may skew the results of the experiment.
* **Corrections**
  + Use a more diverse sample to better represent the galactic population. Maybe include a few wookies, some Humans and some jawas.
  + Increase the sample size to the largest possible size. This likely isn’t difficult to do when you are a galactic sovereign.
  + It would be better to have the speech delivered by the same person to both groups.
  + Maybe have a storm trooper deliver the lines to both groups as they are much less intimidating.

1. In the past, the Jedi have had difficulty with public relations. They send two envoys, Jar Jar Binks and Mace Windu, to four friendly and four unfriendly planets respectively, with the goal of promoting favorable feelings toward the Jedi. Upon their return, the envoys learn that Jar Jar was much more effective than Windu: Over 75% of the people surveyed said their attitudes had become more favorable after speaking with Jar Jar, while only 65% said their attitudes had become more favorable after speaking with Windu. This makes Windu angry, because he is sure that he had a better success rate than Jar Jar on every planet. The Jedi choose Jar Jar to be their representative in the future.

* **Flaws**
  + Each candidate was sent to four planets, however the pre-existing attitudes of the planets they were sent to was different. This means the context of each study was different.
* **Corrections**
  + A better approach would be to have each candidate go to two planets that felt favorably towards the Jedi and two planets that felt unfriendly towards the Jedi. This would reduce the bias present at the start of the experiment and lead to a more comparable situation.

1. A company with work sites in five different countries has sent you data on employee satisfaction rates for workers in Human Resources and workers in Information Technology. Most HR workers are concentrated in three of the countries, while IT workers are equally distributed across worksites. The company requests a report on satisfaction for each job type. You calculate average job satisfaction for HR and for IT and present the report.

* **Flaws**
  + We cannot do a direct comparison of the two jobs because there will be bias introduced by the two countries in which there are IT workers but not HR workers. By calculating an average job satisfaction we are introducing this bias through these two lurking variables.
* **Correction**
  + We can do a combative analysis of the job satisfaction of these two roles in a more limited context. We can do a comparison of the two roles by discounting the data from the two countries that have IT workers but not HR workers. Otherwise we will be facing Simpson’s paradox.

1. When people install the Happy Days Fitness Tracker app, they are asked to "opt in" to a data collection scheme where their level of physical activity data is automatically sent to the company for product research purposes. During your interview with the company, they tell you that the app is very effective because after installing the app, the data show that people's activity levels rise steadily.

.

* **Flaws**
  + By asking users to opt-in we are excluding the users that do not opt-in. This limits our sample.
    - There may be an inherent non-random bias between the users that do opt in and do not. For example more athletic people may be more likely to opt in.
  + The data needs to be looked at over a long period of time. Many users install an app, use it for some time, then get bored and stop paying attention to it after some time.
* **Correction**
  + Removing the choice of opting-in would improve the sample gathered by the app, although this would be a less transparent and more invasive way of collecting data about the users.
  + By increasing the time over which the data is collected we will be able to see if the users are actually improving their fitness or whether they are just playing with a new app in the early stages after installation.

1. To prevent cheating, a teacher writes three versions of a test. She stacks the three versions together, first all copies of Version A, then all copies of Version B, then all copies of Version C. As students arrive for the exam, each student takes a test. When grading the test, the teacher finds that students who took Version B scored higher than students who took either Version A or Version C. She concludes from this that Version B is easier, and discards it.

* Flaws
  + If all the students that take version B are sitting next to each other than it will be easier for these students to cheat compared to the students that have versions A or C.
* Correction
  + The tests should be randomly mixed up prior to distribution rather than being stacked.