Challenge – Evaluate an experimental analysis

1. The first issue I have with this part is that the sample groups were given different presenters. Emporer Palpatine spent years in the galactic senate and is well versed in the art of public speaking and manipulation. Darth Vader(once a promising Jedi in training) is widely known to let his anger and emotions control him and would therefore present even the same slogan in a very different way than palpatine.
2. If Windu is sure that he beat Jar Jar on most planets then it could be a case of simpson’s paradox and the size of the populations that Windu and Jar Jar faced were vastly different. But the main culprit here is that Windu and Jar Jar were given the same task with very different testing conditions. Friendly planets are much more likely to listen to a speaker and their ideas(even if that speaker is a terrible plot development only put in for kids to like and buy merchandise from Lucas) while Mace Windu is not exactly known as an expert negotiator of hostile audiences(this is the guy that got finger-lightninged out a window after failing to convince Palpatine to surrender).
3. Two major issues with the sampling/testing here. The first being different jobs have much different satisfaction rates. While I do not have a lot of experience with IT people, I know a lot about HR and it is one of the most stressful and highest turnover prone positions at any company. That alone should mean that these positions cannot be compared. Secondly, job satisfaction varies greatly from country to country because of different cultural norms for working. Swiss workers for example would expect much more in terms of vacation time, benefits, pay, etc than chinese workers would and their corresponding ‘job satisfaction’ scores would show it. Instead of averaging across the company, the company should compare their individual departments to similar departments in that country and see how they are doing. Aka ‘Are our IT people in Italy as satisfied as other IT people in Italy? And are our HR people in Chile as satisfied as other HR people in Chile?’
4. First and foremost…probably shouldn’t call it a ‘scheme’. Second, it would seem like the app is a success if it gets people to increase their activity levels. What would be good to test here is if those levels are sustained over a significant period of time. It could be compared to gyms where memberships in the first month or two of the year spike because people say ‘I’m going to get fit this year!’ and then they slowly stop showing up as the year drags on.
5. Obviously you would need to randomize the samples. Kids who show up early (maybe they are over achievers) are all going to get version A while kids who show up late(might not be as concerned with the class) get version C. Secondly the fact that a version had higher scores does not necessarily mean it was easier, rather it could have presented the problems in a better, more understandable way that was confusing to takers of Tests A and C.