Experiment Design

1. What **assumption** are you testing?

2. Describe the experiment **stimulus** and / or **data** you plan to collect.

3. Who is participating in this experiment?

a. Describe the participants	**b. How many are participating?**

How will you **decide** if this hypothesis passed or failed?

4. Metric		**c. Threshold**	**d. When**
a. Action	b. People		

5. Why will the stimulus impact this metric?

6. After you run this experiment, what **action** will you take as a result of the data that you collect?

If it **Passes**	
If it **Fails**	
If it is **Flat**	

Experiment Design Template Instructions:

1. Start by identifying your **riskiest assumption** from the Evaluate Evidence worksheet. Enter that into box 1.
2. For the assumption that you identified, ask yourself, **If this assumption were true, what would you expect to see in the world as a result?** Think about either a) an **experiment** you could run to determine if that is true or b) existing **data** that you could collect to evaluate if that were true. Describe it in box 2. Use the following example as guides:
    a. Suppose I'm working on a new feature that helps people manage their taxes. My idea might hinge on the assumption that, "People know how to select a retirement plan." If I ask myself, what would I expect to see if this assumption were true, I might answer, "I would expect that if I talk to people about their process for selecting a retirement plan, I would expect that most of them will express some level of comfort with doing so."
    b. My idea might also hinge on the assumption that "People want help filing their taxes." If I ask myself, if this were true, what would I expect to see, I might answer, "I would expect to see that some people are already getting help with filing their taxes." The data I would want to collect to support this assumption might be the % of people who use a tax professional like H&R Block, TurboTax, or other tax preparation software.
3. Now think about **who are the best people to participate** in this experiment or **whose data you will use** to evaluate your assumptions. In box 3.a. describe the qualifying criteria for your participants. In box 3.b. enter **how many people** will be participating.
    a. Continuing with our previous examples, I might interview people in their 30s who have already selected a retirement plan and have indicated they want help managing their taxes. I might plan to interview at least 8 people.
    b. For the data collection example, I might use the same audience if I have a way of filtering the data according to this criteria. If I don't, I might have to approximate the best I can—perhaps people in their 30s. If I can't get industry-wide data, I might decide I need to track down data for at least 1,000 people for me to have confidence in the data set.
4. Next **decide how you'll decide what the data means**. Define what **metric** you'll use to evaluate whether your hypothesis is true or false and enter it in box 4a or 4b depending upon what you are counting—actions (x emails opened) or people (y people opening email). It's okay to have more than one metric, but make sure that each metric impacts your decision behavior (more on that in step 8).
    a. My metric might be the number of people who indicate they are comfortable with selecting a retirement plan, or it might be the number of times a story was shared where people indicated they were comfortable selecting retirements plans (since people can select a retirement plan more than one time), or I might design a comfort scale and use an average comfortability rating as my metric.
    b. In the second example, I might count the number of people who use tax software.
5. For each metric, you need to define a **threshold** that your metric must clear in order for your hypothesis to pass. This threshold can be **absolute** (30% open rates) or **relative** (10% improvement). Enter the threshold in 4.c.
6. For each metric, you also need to define when you will take your measurement. Enter it in box 4.d.
    a. In our first example where we are interviewing people, we will take our measurement after we've interviewed all of our participants. For some experiments, you might need to build in a lag. For example, if you are measuring email open rates, you don't want to take your measurement right after you send the email as you want to allow time for

people to open the email. But you do want to define up front when you'll take the measurement—maybe 5 days after the email was sent.

   b. In our second example, we need to define the bounds of the data we will look at. Are we going to look at the lifetime of each participant or just the last five tax years?

7. In box 5, you want to describe **why** you think the experiment stimulus or data mining will impact the metric as you outlined. You want to make sure there is always a strong reason connecting the stimulus / data set to the metric as this will help guard against spurious correlations.
8. In box 6, you want to decide up front what decisions you will make if this hypothesis passes, fails, or is flat.
   a. **If a hypothesis passes**, you typically want to move on to testing your other risky assumption. Or if this is the last risky assumption to test, your next step might be to build the feature.
   b. **If a hypothesis fails**, you need to be careful about what conclusions you draw. Revisit the Generate Assumptions worksheet.
      i. If the assumption you are testing is a **design assumption**, you probably want to test other designs before you give up on the feature or the value proposition. If you've tested multiple designs and none of them are passing, then you might need to consider other features.
      ii. If the assumption you are testing is a **feature assumption**, you probably want to test other features before you give up on the opportunity or value proposition. But if you've tested multiple features and none of them are passing, then you might have to consider other value propositions.
      iii. If the assumption you are testing is a **value proposition assumption**, then you might need to consider a new value proposition.
      iv. If the assumption you are testing is a **feasibility assumption**, then you might need to reconsider your feature / design choices to get to a more feasible solution.
   c. For split tests where you are comparing a new design to an old design, you'll want to consider what you'll do **if the hypothesis is flat**. That means that your new design performs the same as your old design. In some cases, you might want to keep the new design for reasons other than performance. In other cases, you might want to continue to search for and test better designs.