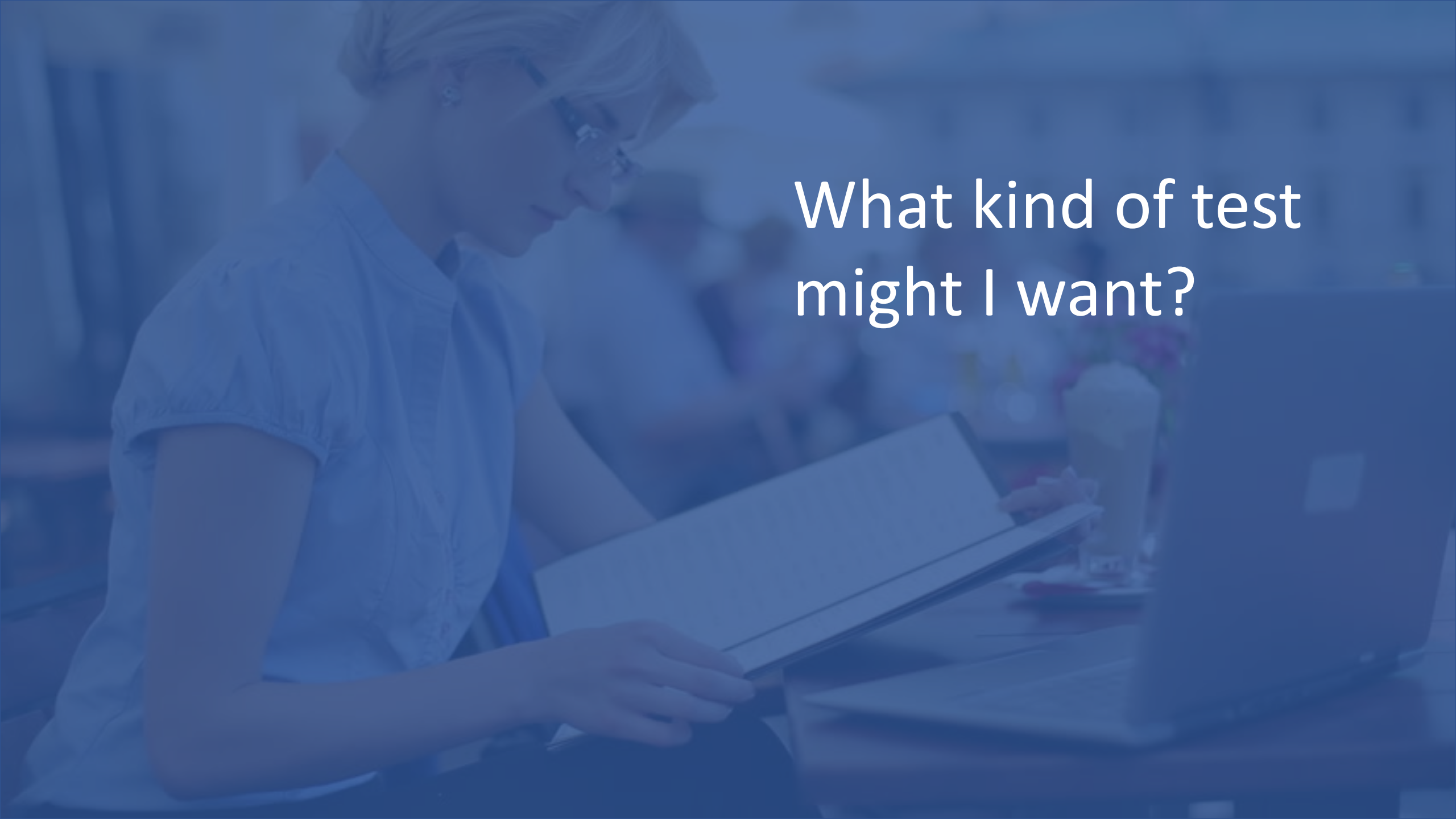
Should it Test?

A guide for working with your researcher



What kind of test might I want?

For tactical we can measure usability such as…

- Efficiency
- Learnability
- Memorability
- How many errors do customers make *
- How well do customers recover from those errors
- Customer Preference **

Asking for customers' preference

May yield useful data

- Flows
- Features
- Holistic experiences
- Costs (what price people would buy it at – requires a survey and large sample)
- Content (what makes sense)

Will not provide useful data

- Color of buttons
- UI Elements (customers aren't designers – even if we use participatory design this shouldn't be our takeaway)
- Position outside of saliency or usability

For tactical we can also

Pain Points	Find usability issues and pain points with competitors sites/apps
Features	See how customers interact with features Do they find these useful?
Discover	Test navigation and discoverability Get a sense of what features are the most salient
Time & Errors	Time on task + error rates (quantitative evaluations of high fidelity prototypes or builds)

For generative we can

Find unfulfilled needs customers have

Create a “Jobs to Be Done” framework about what drives customers’ behavior

Find out customer perceptions of a technology or company

Examine customer buying patterns and strategies

Figure out customer mental models

Dive deeper into current customer pain points with a process

And many more things – bring me your questions about **why** or **how** customers do things and we can find a method.

Generative vs. Tactical

- It's valuable to identify when research questions are generative and when they are tactical
- Helps to communicate capabilities, limitations, and set expectations among team members
  - Speaks to which methodologies are most appropriate
  - Addresses timeline and resource requirements

	Directives	Methodologies	Patterns of Effort	Output Artifacts	Timeline and Sample Size
Generative	**Exploratory** in nature, **abstract**, mapping cognitive strategies	**Less Control/Participant Led** • Contextual inquiry • Interview/shop-along • Diary studies • Social cont. text anal.	Comparatively **simple protocol design** and **effortful data analysis**	**Interpret motivations/behaviors** • Reports • Journey map • Personas • **Jobs Atlas**	**Longer timelines** (~ 8 to 12 weeks) **Small sample size** (n ~ 5 to 20)
Tactical	More direct, **concrete**, getting **feedback on specific test stimuli**	**More control/Researcher Led** • Usability studies • Surveys • Card sorts • A/B tests • Eye Tracking	Comparatively **difficult protocol design** and **simpler data analysis**	**Summary of observed responses** • Tables • Notes + Screen Shots	**Shorter timelines** (~ 1 to 2 weeks) **More variability in** sample size (n ~ 5 to 100+)

Research we have the resources to conduct

- User Testing (remote)
- In person testing
- Qualtrics (Surveys)
- Participatory Design
- Bulletin Boards
- In Store Shop-alongs
- In Home Interviews
- Diary Studies
- Tree Testing or Card Sort

Method	Pros	Cons	What We Can Learn	Example Uses
Remote Testing	Cheap, fast, requires much less advance notice for scheduling than any other method.	Some technical constraints.	What people say and how people behave with a piece of technology.	Usability Tests Contextual Inquiry
In Person Testing	Less technical constraints than remote testing	Expensive (at least $130 a person) and requires a week advance notice to schedule participants	What people say and how people behave with a piece of technology.	Build tests
Surveys	Fast and able to get a large enough sample to have confidence in it's representation.	Data is less rich – has less whys and context.	Customer preferences How patterns seen in qualitative methods look in a larger population Card sorts	Very tactical and straight forward questions. Or used in tandem with a qualitative method to see if those patterns carry over to the population at large

Method	Pros	Cons	What We Can Learn	Example Uses
Bulletin Boards	An inexpensive way to gather longitudinal data. Can use a variety of multimedia. Customers are able to communicate with each other and it will influence how they use things.	Customers are able to communicate with each other and it will influence how they use things. Lots of data -> longer analysis time	Good for evaluating complex purchasing decisions or home projects over a longer period of time.	How do people decide what to get? How do people tackle their home projects? How do people use how-to-content?
Participatory Design	A way to get deep insights out of customers. Creates a sense of ownership and investment in ideas beyond just expressing opinions.	Large logistical lift to implement well. Data analysis takes longer than some methods.	What is important to users. Users ideal product	When we are pie in the skying new features or products. When we want to understand what features are most important to users. There has to be a bit of flexibility with the scope/ design of the project

Method	Pros	Cons	What We Can Learn	Example Uses
Dairy Studies	Longitudinal data Users are in their native environment rather than a test situation Inexpensive compared to other longitudinal methods.	High drop off rate Not able to probe deeper into customers thoughts and opinions as it happens (and they might not remember later) Limited to customer thoughts rather than behaviors (self reported) Very time consuming analysis	Habits Attitudes and motivations Customer journeys	Complex purchasing Projects Creating a journey map
Card Sort Or Tree Test (reverse card sort)	Simple, cheap, fast. Can yield a lot of information with little effort. Scales well to high sample sizes	Lacks context and “whys” People can be inconsistent with how they sort/group things based on cognitive biases.	Information architecture Customer mental groupings	Redesigning/ recategorizing the filters menu Navigation menus

Method	Pros	Cons	What We Can Learn	Example Uses
In Store Shopalong	Very true to life context with variables that would be difficult to replicate. Can create circumstances for a purchase to occur.	Expensive and requires a lot of planning. Also difficulty with quality of audio recording.	How people shop for ___. How people use applications in store	Shopping for or with whatever you are wanting to learn about.
Eye Tracking	Provides additional information on user behavior that more traditional methods don't give.	Isn't exact (they have to be looking at something for longer than a certain threshold). Is mostly good at measuring saliency. We have to figure out how to get the gear (I think EUX has some) and do set up.	Discoverability of features Saliency of features For features that aren't being used: are people seeing them and not engaging with them, or are they not seeing them?	In store shopping experiences

Method	Pros	Cons	What We Can Learn	Example Uses
Observational Research	Natural environment True to life behavior Able to observe more contexts and group dynamics than created test environments.	Are not able to find out the “why” of behavior or motivations behind the behavior without disrupting the flow.	How people really behave doing ___	In store or on construction site behavior
Jobs to be Done	Broad and deep insights into customer motivations	Very time consuming	What drives customer behavior Success factors Current pain points	Most tasks that have participant goals or activities

Method	Pros	Cons	What We Can Learn	Example Uses
RITE Testing (Rapid Iterative Testing and Evaluation)	Fast Team involvement	Requires a lot of participation from designers (similar to a design sprint in lift) The research will have actionable, but not deep insights.	Usability issues and customer preferences	Early in a project lifecycle
Ethnography	Natural environment Deep insights Participants may confide and share thoughts they wouldn't in more superficial research methods	Time consuming and expensive Dependent on researcher relationship with participants. Difficult to recruit	Why people do what they do	Finding out a target market's needs

How To Get Started for a Moderated Test

The product owner view

Kickoff Meeting

Set up a kickoff meeting so we can get on the same page and I understand your test objectives and goals.

Study Plan Review

I'll send out a study plan to review with you and ensure we are gathering the type of data you want

Testing

If your prototype is being tested, I expect you to attend at least 2 sessions to see customers interact with it. If we are testing a competitors site it is just a nice to have.

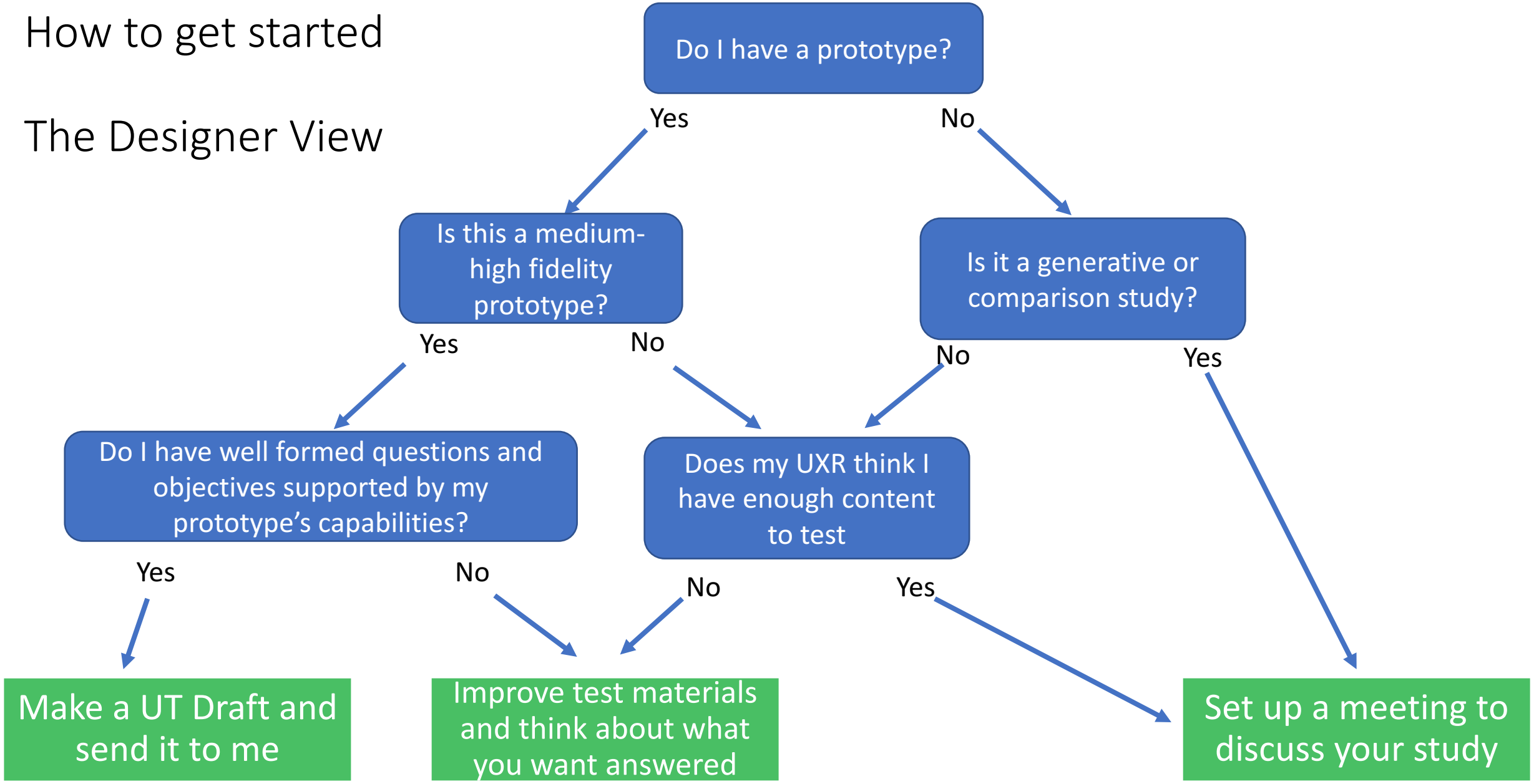
If you attend the first or second session, you can see the type of information coming out of the study and we can discuss changes if needed

Topline Report

48-72 hours after the conclusion of the study, we'll meet regarding the high level findings from the study. Then we will determine the deliverable that matches your needs for the type of data discussed – if we didn't decide on one in the study plan review.

How to get started

The Designer View



TIMELINE
~8 weeks

PROCESS

Prioritize | Kickoff | Plan | Research | Analyze | Report

STEPS

- Identify the most impactful research opportunities across communities
- Choose next project
- Hold kickoff meeting
- Explain questions about customer behavior
- Clarify stakeholder objectives into research questions
- Outline specific research objectives
- Determine methodology, customer profile, and timeline
- Provide feeback on study plan and ensure alignment
- Recruit participants
- Interview customers
- Attend customer interviews and take notes
- Aggregate data and observations
- Identify patterns of behavior, motivation, friction, and desire
- Draw conclusions and collect actionable insights
- Compile findings into report deck, visualizations, or other suitable formats
- Present findings to stakeholders and larger UX community

PEOPLE

- Researchers
- Product Managers
- UX Designers
- Engineers

Timelines

Reviewing and Launching a Test– 1 day
You watch the videos

Launching a survey and providing a topline – 3 days

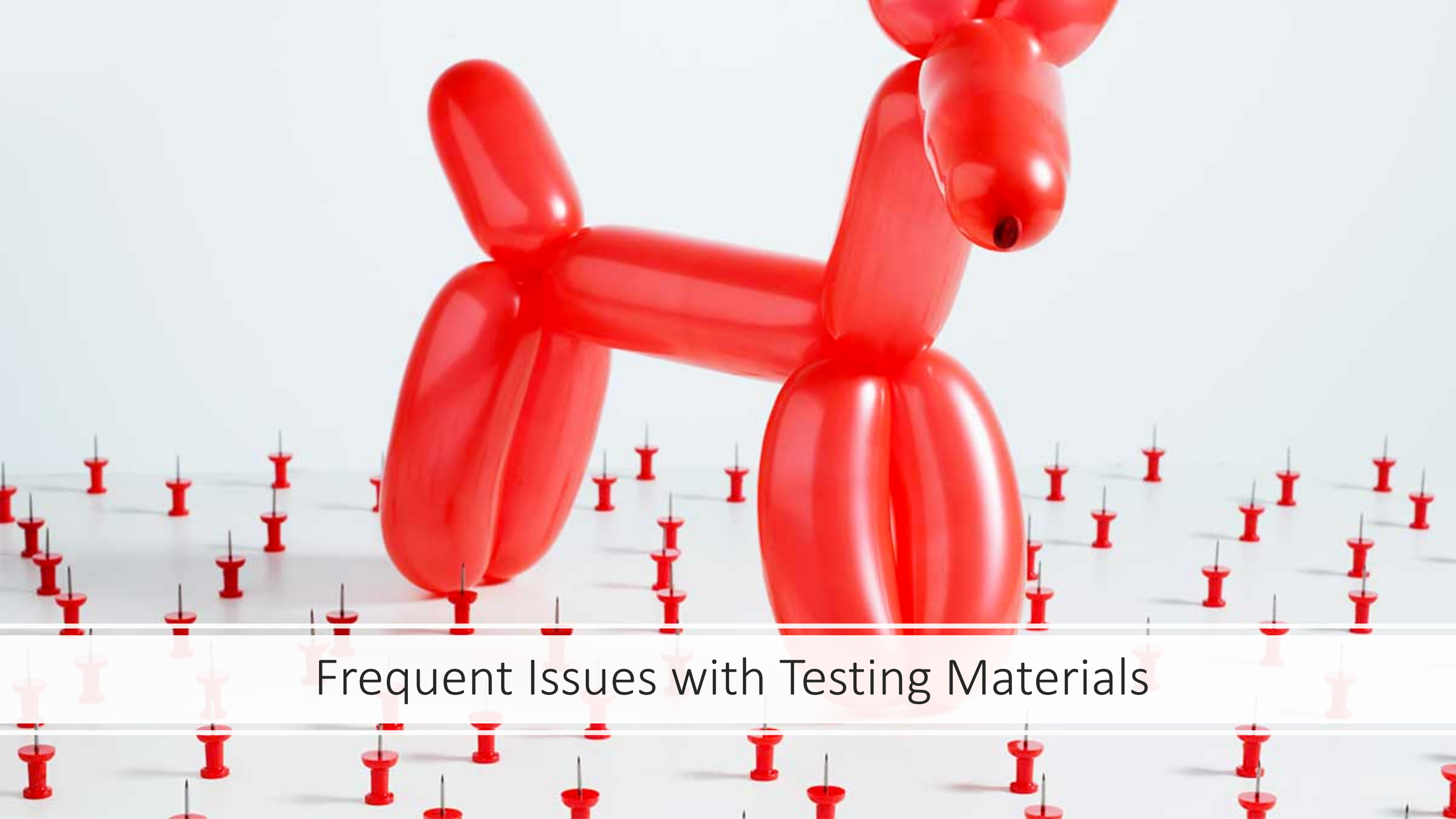
Usability test unmoderated – 1 week

Usability test moderated – 2 weeks (may be sooner depending on ask and scope)

Tactical test (depending on the ask and scope) – 1-3 weeks

Generative test (depending on the ask) – 3-8 weeks

Style of test we haven’t done yet - TBD



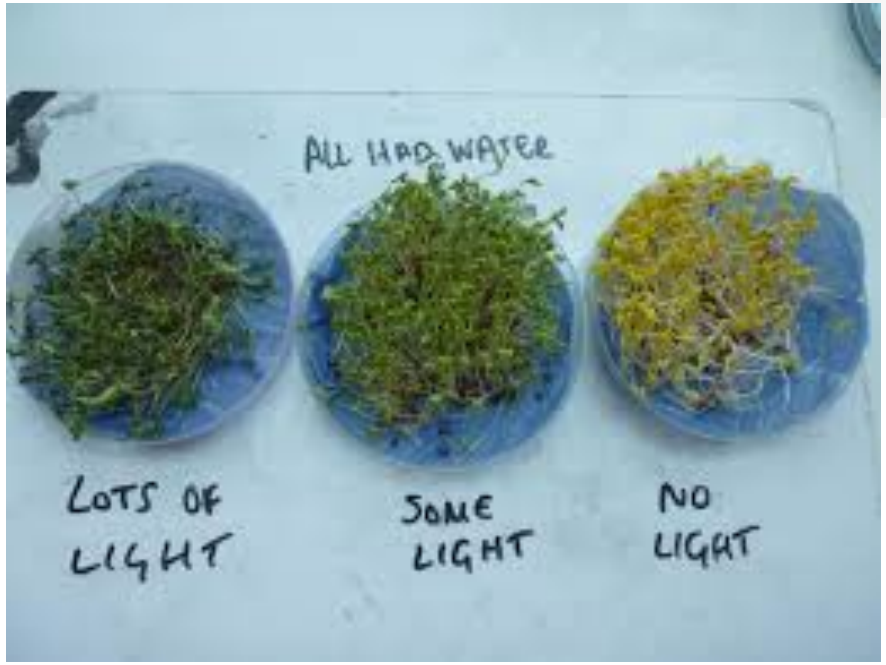
Frequent Issues with Testing Materials

Does it have more than one variable?

- No light
- No water
- No fertilizer
- Not even the same type of plant

- Some light
- Some water
- Some fertilizer
- Species that flowers

VS



Does the prototype interaction support what you are measuring?

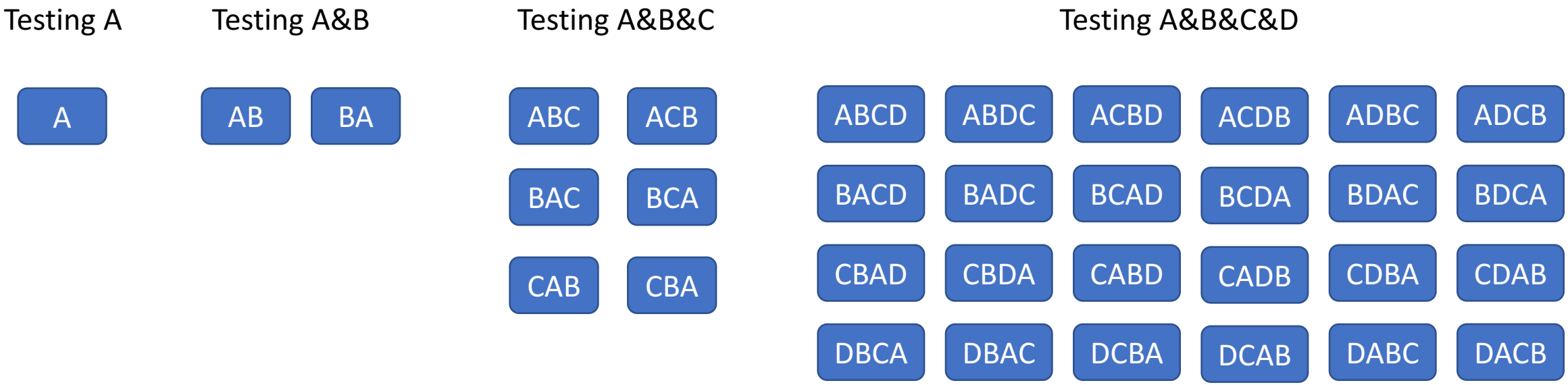
Do you have the complete flow for the specific feature you are testing?

If you are comparing a flow, are both flows equally complete/robust?

Can you questions be answered by customer behavior and opinions about the content

Is the result from the test going to affect the design you launch?

Can we narrow it down to 2 or 3 versions to compare?



Questions?