



Webinar on Implementing Online Report Cards: Research-Based Practices Workshop, Process

Friday, June 15, 2018

2:00–5:00 p.m. Eastern Time

Webinar transcript

[Steve Glazerman]: You need stakeholder input, so how do you get stakeholder input? What-you tell me, what are the different methods that you've either used or feel like are important? Surveys, okay.

[Audience member]: Focus groups.

[Steve Glazerman]: Focus groups. Don't make Jessica to do all the work.

[Audience member]: Task force.

[Steve Glazerman]: Okay. Sorry, I heard nothing.

[Audience member]: Community forums.

[Steve Glazerman]: Community forums, okay. So let's see community forums, surveys, focus groups.

[Audience member]: PTA meetings.

[Steve Glazerman]: Yeah. Okay. Those are types of meetings of the public. Anybody have any other ideas for ways to solicit input from users on how your tool's working?

[Audience member]: Well, we just use like school list serves to get the news out about their behavior, so survey or –

[Steve Glazerman]: Okay. And you can think of those as almost a kind of survey, that maybe not be closed ended questions, but there's also something called cognitive interviews, which we'll talk about what that is. There's a specific type of interview. You can also gather website analytics. If you have something online, you know, if you have a- if you've already launched, you can get a lot of feedback on your site. I should mention stakeholder input, I think I

mentioned this earlier, it's not just something that you do in the discovery phase, you know, like Dave was mentioning at first phase, but you want to keep doing it throughout the lifecycle of the report card. And there's another category that we're going to call experiments or A/B testing. So let me just sort of run you through what each of these is. Alright, so first question you want to think about when to get stakeholder input. It's not just at the beginning, you want to get stakeholder input when you're testing prototypes. You may want to get a stakeholder input after you've launched either because you want to, you're usually going to be updated. And so for next year, you may want to revisit.

So let's start with surveys. It's worthwhile thinking through explicitly some of these considerations when you're about to mount a survey. I mean, a lot of people think, "Oh, I'll get a survey monkey account or some other, you know, low cost or free tool, you know, survey platform. Write a few questions and then sort of send it to everybody I know and everybody can then put on the list serve or post on the- make all of my employees put it on their signature page or maybe we have a blast email list that we use for something. The problem with most of those approaches is that you're going to get a selected sample, you're not going to get a representative sample of the population of interest. You have to go back to that card, you know, that you wrote, who is your population and make sure you're hitting that population. And then you want to be representative of that population. The most obvious way to be representative is to list them all out and draw a random sample. That's most of time not practical, okay? But that is- start with the ideal and say, "How can we approximate that"? You can still do what's called a nonprobability sample. That's anything where you just, you know, circulate a link. Non-probability samples are not necessarily bad, but they can very likely to give you a non-statistically representative sample.

One way to address that limitation is to include questions about demographics or something that you can then take, that you already know that let's say the demographic makeup of your target population or the geographic makeup. So tell me, you know, what city ward you live in or what school district you live in. That way, if you get, you know, over representation from Anne Arundel County under representation from, you know, Baltimore City, you can adjust the responses statistically to give more weight to those who are underrepresented to get you closer to representative sample. And you can do that with demographic groups or anything that you know the overall population characteristics of. There's a trade-off between the quality and the quantity of data. So that's why some people say, "I just need- I'm going to keep going until I get 2,000 responses". I'd much rather have 500 responses that are statistically representative of my population than 10,000 that are not. But it's a trade-off between bias and precision. I'm going to- in interest of time, I'm not going to deliver a lecture on survey design principles that we could do a workshop on that because that's

something that, you know, Mathematica does quite a bit on a national scale, it is designed surveys.

But there's some points that you probably know but are worth reinforcing. Avoid social desirability and question wording. If you're trying to get feedback on your report card, just be aware of, you know, your respondents are going to have- you may be getting your respondents' attitudes about something related to what you're asking about and not the actual response to what you're asking.

Just a couple words about focus groups and cognitive interviews. Everybody- I think a lot of people are very familiar with focus groups but I think people get- they think of them as a substitute for a survey. And it's not a survey, there're some things that a focus group is really good at. A focus group is really good at surfacing ideas, answering 'why' questions. But anytime you're summarizing the results of your focus group by saying X percent of focus group participants said, "Why?" You're doing it wrong because the whole point is it's a group dynamic. You want to motivate a group dynamic so what Will says in the focus group inspires Suzanne to say something that's related. And then it's sort of a dialogue. And then Theodora comes up with, "Oh, that reminds you of that". And then people are nodding their heads. So if somebody said something, you're not recording how many people nodded their heads and how many people scoffed to get the true metrics. So even within that focus group, you don't know what percentage of participants agreed with the statement.

But you heard the statements like, "Oh, I had no idea that it was so important to get the phone number of the school principal". So you learn what's possible. Cognitive interview is something we do a lot when we develop surveys, at least at Mathematica, and it's a little bit more intensive and expensive so it might be less frequently used but it really gives you a lot of insights. It's less about the opinion that a user has about your tool than understanding the thought process. And this is really important to identify the right participant for a cognitive interview because everybody around the room we went through this exercise a moment ago where we looked at the Illinois State report card and the kinds of observations that Josh Boots is going to make about a school report card and his fascination with the scatter plot are not going to really tell you anything about somebody a mile away from where we are right now or even across the street who regular DC resident parent trying to make sense of the school finder that they're going to use to place their kid in the school. We need to get down into the right community, with the right people. And then the cognitive interview is basically a think aloud. If somebody has the artifact, they usually have a task like, determine whether Springfield Elementary School is good for your kid. But they think aloud as they go through it and that could be very valuable as a insight generating device.

And then website analytics is sort of the opposite extreme and that it's cheap, it's like practically free. If you've launched your website it's very easy to get analytic result and this is in some slides that Dave put together, so I'll just show them to you because I sort of come up in this context which is looking for a large state report, some anonymous large states report cards. They looked at the hit statistics and found that most of the usage happens in the first week it's launched, that's a pretty useful insight. So you think you have this tool and it's basically- that's why the newspaper headline- the newspaper story metaphor is it motivates that because it's like it's not what would be what we the other metaphor so that you keep going. And now the dashboard, almost 40% of the users in this application were looking on their phones. Most web analytics tools will tell you something about the device that was used to view the page. And so probably most of you won't be surprised at the rates of mobile usage, but that helps you validate decisions to focus on mobile readability and tablets obviously somewhere in between. This is interesting, most of users just look at few pages. So if you're putting information three, four, five clicks deep, be aware that nobody's gonna know it.

One of my colleagues and I, we did the study of school choosers in DC using the rankings information. We tried to reconstruct all the information that parents had access to and use all those as predictors. One of the predictors was the growth percentile of the school's sort of growth score. Guess what, you how many clicks it took to find the growth score of a school? It was a lot, it wasn't on the main page, it wasn't on the- you had to go to LearnDC and then you had the equity reports or you go to a different site and then you're- and even within the same report card system, the users in this state were not going very deep. This is probably one of the most useful graphs I think, which is how long does somebody spend. These labels are in seconds. So this is how long people are spending on the site. I would actually look at the conditional distribution just this part and ignore that for a second and just note that the modal categories. There's still a three- the modal category, if you ignore that top bar, is 3 to 10 minutes. So it's not like nobody's doing that. But the point is not what is true in this particular state, but that this is free information that you should definitely mine and there's more sophisticated tools.

The last mode that we're talking about is A/B testing, some people call it experiment, rapid cycle experimentation. But the basic idea is you have some kind of stimulus, version A or version B of your site. Or you could just vary one thing, like you have- it may be that you said a lot of things through your community meetings, through your focus groups, through your surveys, you run your cognitive but there's something that you just can't resolve. You're just not sure what's gonna- and it's pretty high stakes like letter grade of the header. So maybe make a version with and without and you randomly assign users to one or the other, measure some outcome of

interest. If you're measuring click through that's the easiest, but you may want to have a survey of knowledge or attitudes. And then just test which version produced the greatest response, it's that simple. It doesn't have to be a very large sample, there's methods to determine how large the sample has to be based on what kind of- have firm a conclusion you need to draw. But this is like basic, running basic science, just run an experiment. And this can be sort of generalized to more than just testing one thing, you can test multiple things at once, we call these factorial experiments.

The study on which this guide is based, the one that I mentioned before, was a factorial experiment. There's actually more things on this page that are worth talking about for this workshop, but the main thing is that, that this is available to you as a methodology that's, you know, pretty straightforward really. But surprisingly few people want to commit to doing it because, I don't know, it's just there's- some things are just more- like surveys or focus groups are more comfortable, right? But the truth is like, even if you have a somewhat nonrepresentative, not a statistically representative sample, within that sample you will get a very good sense if you can randomly assign the different versions to one group or the other, you'll know for that group, which those groups are going to be at least within the total sample balanced. And so we ran this experiment. This is a mockup of a website that we had Tembo designed for us, but we had a map, there's a map up there and there's- each row is a school. This was for school choice so, you want to see all the schools at once. And we tested the format whether you had, this one has graphs, we also tested icons or just plain numbers, whether you have a district average, so this top bar shows a district average. There could be some kind of anchoring effects where having a district average makes you interpret the metric differently because now you have a basis of comparison. If the overall proficiency rate is this is probably the 60s and your school is 87, that looks a lot better than 87 on its own maybe.

We tested whether the data was coming from a survey of parents or some objective indicator like number of suspensions. We also tested the amount of information, so this row right here is a row for one school and then there's a little plus icon if you click it, it opens up the drawer and shows all this detail. So this one has this school, Ward Elementary School has the detail shown. So that's something we tested this progressive disclosure, there's another version of this, same exact website, but you can't open the drawer. It just has the summary. There's another version where the drawer is open for all the schools. So you got about one and a half, two schools on the screen and you scroll down to see more. We also tested this other factor, which is the default sort order. You notice up here, there's a little drop down to sort by another attribute. And you can change it, user can change it, but the thing that we tested was the default. There's a lot of literature, research literature on the importance of defaults. For things like organ donation, organ donations jumped when the default was changed on your driver's

license from no to yes. And you can always change it, that's the whole point. It's voluntary but the fact that it's a default they're very, very sticky. Even in matters as personal as organ donation it could be very influential in the- and then the actual order could then influence your choices.

I'll have to leave you in suspense about what the findings were. But the main point is not about this particular experiment, but the idea that you can run an experiment, test all these factors. We actually have one that we're proposing to do with OSSE where we're going to run it. We're going to let OSSE define what these factors are. We're going to let the developer, the designer implement the variations and then we're going to identify a sample of DC parents. Maybe it's going to be low income parents, maybe we're going to use non-English, I don't know what the intention is and then we're going to randomly assign them to one of these and run this. Basically, use the same sort of outcome metrics that we used in the IES study.

[End of Audio – 00:15:46]

This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006, with REL Mid-Atlantic, administered by Mathematica Policy Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.