

QUALITATIVE ANALYSIS:

Webinar for Teen Pregnancy Prevention 2018 Grantees

Monday, March 9, 2020

1:00 to 2:30 p.m. eastern time

Webinar transcript

Webinar producer: Hello, everyone, and thank you for attending today's event. Before we begin, we'd like to cover a few housekeeping items. At the bottom of your audience console are multiple applications which you can use. You can expand each widget by clicking on the maximize icon at top right of the widget by dragging the bottom right on the widget panel. If you have a question during the webcast, you can click on the Q&A widget at the bottom to submit your questions. If you have any technical difficulties, please click on the help widget; it's a question mark icon that covers common technical issues. However, you can also submit technical issues via the Q&A widget. An on-demand copy of this webcast will be available approximately one day after the webcast using the same audience link sent to you earlier. Also, the recording and materials will be posted next week on the max.gov website. I'd like to turn it over to Diana McCullum. Diana, you may now have the floor.

Diana McCallum: Hi, everyone. Welcome to our webinar today. We're really excited to have a wide range of participants with us that are folks from across the TPP [teen pregnancy prevention] grantee cohorts, and I'd like to start by introducing you to our presenter, Dr. Jane Choi. Jane is a researcher at Mathematica with over six years of experience in survey research and overseeing qualitative analysis and providing evaluation technical assistance. Many of you, I know, have direct experience with Jane. She's an evaluation TA [technical assistance] liaison for seven of the TPP 18 grantees as well as some TPP 19 grantees. She's also provided technical assistance to grantees in other areas, including those developing education, fatherhood, and healthy marriage programs, and she's working with grantees that are developing prescriptive and rigorous evaluations. We're really excited for this presentation today, and to have Jane especially, because she has extensive experience conducting qualitative data analysis. She's led efforts to create qualitative coding tools and conduct analysis and has authored reports that focus on finding some qualitative data. So, I will turn it over to Jane to start the discussion.

Jane Choi:

Great, thanks so much, Diana, and yes, welcome everyone to our presentation. Today, we're going to go over how to conduct qualitative analysis in a systematic way. So, this is our agenda for today. First, I'm going to provide an overview of the purpose of qualitative analysis. And then, I'm going to go over the steps to conduct the analysis, and I will sprinkle in plenty of examples. And then, I'll go over some of the general guidelines for reporting your findings, and we'll have a good chunk of time at the end of questions. Then, I'll share some resources with you all.

So, like all analysis, the purpose of qualitative analysis is it's to be used as a tool to answer your research questions by finding patterns in your qualitative data. And qualitative data can come from a range of sources. So many of you are collecting these types of data now, but they can include observations, interviews, focus groups, open-ended responses from surveys or fidelity logs and then, also, any of your curricular materials. So, findings from qualitative data can help evaluators understand the details and nuisances of how programs were implemented.

So, for example, if the program wasn't implemented as intended, you can find out information about, you know, what some of these issues were, like low enrollment and scheduling changes. You can also learn more about why a program was or wasn't able to meet targets or goals. So, for example, if you found participants didn't perform well on a healthy relationship knowledge outcome, your qualitative data like focus groups can show why that might be the case. So, for example, maybe there was a particular lesson that participants didn't respond well to, and maybe some of the stories or examples in the curriculum weren't culturally relevant. Qualitative findings can also provide a more in-depth look at the viewpoints and perspectives of the participants who were involved, and you can understand more information about their experiences. So, I'm going to start us off with a metaphor for qualitative analysis. I think sometimes, the process of qualitative work can feel pretty ambiguous and up in the clouds, and I feel like maybe an on-the-ground metaphor might help make things feel more concrete.

So, conducting qualitative analysis involves similar steps to organizing a large jumble of papers. So, you can kind of see all these big papers floating up in the air. And you want to be able to organize your papers meaningfully so you can efficiently find what you need. So, eventually, all of these papers, you're going to want them to be kind of stacked in these file folders in the file cabinet on the bottom right. You need to organize these papers in a meaningful way so you can find what you need, and you might realize the process is a little bit messy and maybe frustrating because, you know, there isn't really a set organizational scheme. This is all something that you have to develop. Similar to this, with qualitative

analysis, you have this goal of wanting to answer your research questions for a report or for your journal article if you're a TPP 18 grantee, but you probably have a really large quantity of data. It's probably a little bit messy, and you need to figure out how to clean the data, organize them, and make sense of them in a systematic way. So, this process can be a little bit frustrating because it's not always linear and it's not always clear-cut.

But I'm hoping that after today's presentation, you'll have a clear sense of where to begin and some of the next steps that you can take so that you can answer your research questions using your qualitative data. So, I'm going to provide an overview of the different steps to conducting qualitative data, and I'll kind of go through them really quickly at first and then, after this slide, we'll go through them in the step-by-step way with examples. I also want to note that there are multiple steps, and I present them in a way that makes them seem linear, but a lot of this work will kind of happen in tandem and you'll toggle back and forth between the steps. I'll do my best to note where this happens, and then I'll also note it as I go through each step in detail.

Our first step is to refine and focus on our research questions. Before we start anything, we have to start with our questions, because the whole purpose of doing this analysis is to answer these questions. So, you'll want to make sure that you refine your questions so that you have a final version of questions before you start your analysis, and you want to make sure that you understand the purpose of these research questions like why are they important to answer and why are these the questions that you've singled out.

And then, with every subsequent step, you'll want to refer back to your research questions and make sure that everything you're doing is aligning to them. The second step is to get the data ready to be coded. This will involve cleaning the data and making sure that they can be easily understood and that they're ready to be analyzed. The third step is to develop a coding system. You'll do this by identifying codes that align to your research questions, and a code is basically a concept or an idea that you will use to label portions of your data based on what the data—or what the meaning of the data—is. The fourth step is to use the coding system to code your data. Evaluators often toggle back and forth between developing a coding system and coding their data, which is why this arrow is double-sided. The fifth step is to group and regroup coded data by theme and often, you'll find that your codes are really specific. So, you might want to regroup your codes into larger, broader codes by theme so that they'll be more meaningful. And then, the sixth step is to determine what your findings are.

So, after you've coded your data, even after regrouping them into broader codes, you might still end up with a really long list of codes, and you're going to need to figure out what is a finding worth reporting so that you're not reporting, you know, 15 or 16 codes. The seventh step is to report your findings in a way that clearly addresses your research questions and conveys the meaning behind the findings. This includes explaining what the importance and significance of the findings are, and this is what people mean when they say tell the story of your findings, you know, how they're connected to one another and why they're important. So, I'm just going to highlight again that with every step, you're going to want to make sure to go back and think about your research questions and make sure everything you're doing is aligned to the research questions. And so, you'll notice on this figure, the first step with the research questions is a big box that kind of covers all the other smaller boxes because you're always going to be referring back to those research questions. So, walking through the steps—

Diana McCallum: Sorry, if I could just interject before you continue, just the end of that slide would be great. I just wanted to point out that, you know, we know that our audience today includes, in addition to TPP 18 grantees, other grantees that are at various stages. So, we think these steps will be helpful for not only those that are starting to think about journal articles but also, for those of you that are thinking about how you'll set up your qualitative analysis in the future. We encourage you to reference these slides as some of you get closer to looking at your qualitative analysis. Sorry to interrupt. Go ahead, Jane.

Jane Choi: Oh, yeah, thanks so much, great point. I'm going to walk us through an example study. I think that as you've heard with these steps, it can sound really abstract. And so, I think that giving you examples from one made-up study might help because you can see how all the steps are connected and how they all work together. So, we have some research questions for this study. The first one is, what were key challenges the program team faced when implementing the curriculum? So basically, focusing on key implementation challenges. And then, our next two questions are what curricular or implementation modifications were made and did these modifications mitigate or address the implementation challenges? So, really, what we're focusing on here with these three questions is, what were the key program implementation challenges, how were they addressed, and was it addressed in a useful way?

For this presentation, just to keep things concise, I'm really going to focus all the examples on the first research question, but I wanted to give you all three so you can kind of see that the first question really isn't informative

enough for a report on its own. But when you combine them with these other two questions, you have a more well-rounded set of questions.

Our example study also includes some data. Our data comes from focus groups with participants, and in this fake study, these are students at partner organizations. We also have interviews with the frontline staff or the program team, and we call them facilitators. So, now, I'm going to walk us through each of the steps. Our first step is refining and focusing on the research questions, and, like I said, the whole purpose of conducting these analyses is to answer our research questions. And so, before we do anything else, we want to make sure that we have our research questions set.

And this might involve reviewing your original research questions from your application or your evaluation plan and making refinements as needed. TPP 18 grantees, you worked on this when you drafted your abstracts. Then, you'll want to use your research questions to guide your analysis and reporting. So, you need to be pretty familiar with your research questions, know why they're important and know them inside and out. And then, you're going to want to continually refer back to them as you analyze and report your data. So, I know, many of you are experiencing this, but qualitative analysis usually involves just an overwhelming amount of data, and frequently checking back in with your research questions is the best way to ensure that you're being purposeful and you're not getting lost in your data. I just wanted to note that while we recommend that you refine your research questions from any original plans, you shouldn't be continually refining your research questions as you go through the steps of analyzing your data. So, once you've finalized your research questions, they're final, and you should move forward and reference those final versions. Our second step is to get our data ready to be coded.

So, you want to take your raw data and create clean data files. And clean data files can be transcripts or detailed notes. So, your raw data are usually pretty difficult to analyze. Some examples of raw, qualitative data can include rough notes or jottings that you handwrote or typed really quickly while you conducted the interview or focus groups. They can also be audio or video recordings. For example, if you observed a session, you may have video recorded it. It's really hard to analyze those data, and it might be really time consuming. And so, usually, what researchers will do is create clean data files that are in written format. The best person to create clean data files is the person who is involved in collecting the data, because they'll have access to nuanced information on people's tone and will be able to fill in any holes or gaps in the transcript or the notes. And, of course, we recommend creating these files close to when the data were

collected, so, you can still rely on like your fresh memory. This isn't always possible and so, if it is impossible, you know, its good practice to have the person who collected the data at least review the clean data files to make sure things are accurate. And then, you're going to want to make sure to name your prepared data files in a systematic way, and I like to include this descriptive information about what's actually in the file, and I'm going to walk us through an example of this momentarily.

So, our first example will be about preparing data files. This table shows your first the research question at the top. What were the key challenges the program team faced when implementing the curriculum? And then, on the left, we have our raw notes from a facilitator interview. Let's say, with a facilitator named John. And so, on the left, it's not clean yet, so I'll read it really quickly. "Sometimes, the discussion questions are too baby-ish for the students. Although, it may just be content too young cuz the question vocab is too hard sometimes for them. Like, why do you show affection to someone? The kids don't know the word affection but they think, it's silly to explain why you give someone a hug." So, there's a sentence in the middle there where it's pretty hard to understand. It's missing some words and there are some abbreviations. On the right, I have the clean data, and some of the edits that I made are in that kind of burgundy red-ish font. And so, the purpose of cleaning it was to make everything easier to understand.

So, now it says, "Sometimes, the discussion questions are too babyish for the students. Although, it may just be the content of the questions is too young because the question vocabulary is sometimes too hard for them. Like, why do you show affection to someone? The kids don't know the word affection, but they think it's silly to explain why you give someone a hug."

So, by cleaning the data, you can kind of see clearer that, you know, there are these two things that are working together here that the discussion questions content is not age appropriate, it's too young for the students, but the vocabulary is maybe still too advanced. And so, the goal of cleaning the data here was to make the meaning really crystal clear. Our second example is a screenshot of a data filename. And so, like I said earlier, it's nice to have data files that have descriptive information in the name. And that way, you can quickly know what's in the data file without having to open each data file. So, you might end up with, you know, 30 or 50 files, and you don't want to label them all like data one, data two, data three, because then you don't know what's in them.

So, these are some examples of the type of characteristics that I would include in a data filename. I would put site name. Here, we have boy

charter. The respondent type for our case, is a facilitator. The data source, which is an interview, the data, I usually put the year first and then the month numerically so that it sorts chronologically in my file folder. I'll put the name of the month in parenthesis just so I can see it at a quick glance. And then, I'll put whether the data were collected before, in the middle, or after the program. Here we can see it's during the program, it's in the middle of the program. And continuing on with our third step, once we have our clean data and they're named in a systematic way, we would add these files into a database. I recommend using a qualitative software program such as NVivo or Atlas.ti. They have a lot of built in functionality, and one of them is to keep your probably very large data set organized, and then later, there's a lot of coding and analysis functionalities that are really useful.

If you're not able to use a qualitative software program, you can code in Word or Excel but if you go this route, you might have to take some additional steps to make sure that your coding and your documentation is organized. We can discuss this further in the Q&A portion if this is a route that you want to take.

When you're adding files into your database, make sure to only include data sources that will answer the research questions. So, for example, with our fake research study, we have data from students and facilitators and those are relevant because they're talking about program implementation and program receipt. So, we would include those in the data set. But if I had data from community members about the need and demand of the program, I wouldn't include those into the database because they don't focus on program implementation. I wanted to say a quick note, which was that, in this example, we're focusing on one research question, and if you have multiple research questions, you want to make sure that you're including all relevant data for all of those questions.

In other words, you're not going to create a database for each research question. And our next step is you're going to want to develop a coding system by creating codes for each research question. So, a code is a concept or a label that's used to assign meaning to a portion of your data. You're going to use these codes to label or codify the data to group similar data together. Codes are usually pretty detailed and specific. Examples of codes can include difficulties engaging participants or low attendance. You can see with these two codes that they are a little bit more specific. We could've kind of had one code that focused on, you know, difficulties engaging participants and included low attendance there, but I chose to keep them separate so that we have more specific codes. And I recommend having more specific codes because it's more efficient to group data from specific into a broader code later rather than to take data

that are coded under a broader code and then reread each of the pieces of data to determine how to parse out and distinguish them from the other data and apply like more specific codes later.

I prefer to go really specific in group later. But keep in mind that there's no one right way to do this. You can start broader and then parse them out into more specific codes. There are two methods for developing codes, and I recommend using both methods. The first method is creating codes before you start coding. So, you would create codes from your knowledge about prior research, the focus of the research questions, and OPA core themes if they're applicable. You can also use your project director and your frontline staff and have a brainstorming session to come up with some codes that, you know, people have experienced as they were implementing the program or going through the project. So, the first method is to develop codes before you start coding. The second method is to develop codes as you're coding and as you're going through the data. It's really difficult to anticipate all of the codes before reading the data. As you go through the data, you're going to come across new topics or ideas that you didn't think of before, and you should definitely add those in.

So, like this says, you'll develop codes as you read through the data and then, make sure that you're keeping track of when you create new codes so that you can apply them to data that have already been coded. So, for example, let's say you have 30 interview notes, and John's interview is the first one. So, you code John's interview and then you code three or four more interviews, and as you're going through those, you discover these new ideas and so you add them to your coding system. You're going to want to make sure to code the remaining, you know, 20 or so interviews before you go back and recode John's interview. And that way, you're not coding John's interview 10 or 11 times as you develop new codes. You're only going back and recoding his once. And in terms of what the coding system should include, you should include codes that are organized by research questions and at least a clear definition of each code. I recommend also including the date of when you created the code. I'm going to provide an example of a coding system. I create these in Excel, so I'll have a table where I have a research question at the top and then I'll put my codes on the left, a definition in the middle, and then, on the right, I'll note when the code was created.

I can go through a couple examples of these. With some of the codes, you'll notice it has a definition that's pretty straightforward. It matches almost directly to the language of the code. So, for example, with unresponsive participants, which is the second code in the list, the

definition says, participants are not responsive to the curriculum, the lessons, or any other aspects of the program. That's pretty straightforward to the name of the code. But then, with other ones, you're going to need a more concrete definition. If you go to the last code, issues with literacy and text, the definition is that participants weren't interested in the curriculum that required reading and writing, and participants don't want to engage with long text passages. So, you'll kind of see that, issues with literacy and text is a little bit ambiguous without the definition. As you're going through your coding, you'll want to update definitions and make sure you include more examples.

And you want to make sure that the definitions give you a really concrete history and clear depiction of what each of the codes mean. This is good for record keeping and to make sure that everybody's always coding consistently. Also, these are the types of things that you'll want to report in your methods section of your report. The right column has the dates for when the codes were created, so you'll see that the first two codes were created before we started coding and the second two codes were created in mid-January. If there were any interviews that were coded prior to these dates, you'll want to go back to those interviews and apply these codes if they're relevant.

Okay, our next step is applying the codes that we've developed to our data. So, if you have a lot of data, you might need to have multiple coders and if you are using multiple coders, it's critical to have an initial training and then, ongoing training so that all of the coders know how to use the software, they understand the coding system and they're really familiar with the research questions. So, prior to having coders begin, you will hold a training that covers all of these different items.

After the training, you'll have the coders and whoever led the training code the same interview. And then somebody will check and make sure that everybody's using the coding system the same way and applying the codes to the same portions of that interview. If there are any discrepancies, you're going to want to make sure that you re-explain whatever the codes are, where there are discrepancies, and explain how to apply them correctly. You'll probably want to do this exercise again and check again to make sure that everybody's being consistent. And it's good to have regular checks of coding as people are coding as well. After they've coded five or so interviews, you might want to do this exercise again to make sure everybody's still on the same page. You really want to make sure that people are always consistently coding. One of the benefits of having multiple coders is you're able to assess consistency or reliability, and that's something that you'll be able to report in your report or your journal article.

In terms of actually going down to the nitty-gritty of coding the data, the first thing you want to do is do a quick read or a skim of the full data to get a picture of the topics covered. If this is John's interview, you're going to want to skim the full interview. That way, you have a holistic idea of what John was talking about. And then, when you code, you'll really zoom in and read the data carefully and assign codes as you go. So, you could assign codes to even one or two sentences. And, because you're zooming in so closely when you code, that's why it's really important to do that broader scan of this full data. When you're coding, you're going to assign codes that accurately represent each portion of the data, and you might assign a code as a sentence like I said. You can apply multiple codes to the same portion of data if needed. You'll find that in your interviews, even a sentence can contain a lot of different information, and it might need multiple codes to fully capture the meaning of that sentence. And also, the same information in your data can be used to answer multiple research questions. Of course, while you're coding, make sure that you're always going back to the research questions and keeping them in mind so that everything you do is aligned to those questions. Now, I have an example of how we'll code data.

On the left, we have a list of codes and then, on the right, we have John's interview. I added a few sentences at the top. I'll go ahead and read that again. "Some of the kids didn't want to read or be read to but all of the lessons involved reading. Everything is reading or writing, reading or writing. They liked the discussion but not the reading or writing. They really liked the discussion when the questions are right. Sometimes, the discussion questions are too babyish for them. Although, it may be that the content of the questions is too young because the question vocabulary is sometimes too hard. Like, why do you show affection to someone? The kids don't know the word affection, but they think it's silly to explain why you give someone a hug."

So, in this paragraph, you see a few things. In the beginning, John is talking about how some of the students don't really want to engage in the reading or writing aspects of the curriculum, that they like the discussion when the questions are at the right content level, but sometimes the content is too juvenile and sometimes the question vocabulary is too difficult. And then, there's an example of a question from the curriculum.

So, I'll walk us through each of the codes, and we can apply the ones that are applicable. So, the first code is scheduling and timing. In this paragraph, we don't hear anything about scheduling and timing, so, we won't apply that code. The next one is unresponsive participants. While John did say that some of the students don't want to read or be read to, he didn't say anything about participants not responding to this section. So

we won't use this code on this data. The next code is disruptions to sessions. Again, there's nothing about disruptions to sessions, so we won't apply this code. John didn't say anything about late arrivals, he didn't say anything about early departures. Our next code is that the curriculum's not appropriate for the age. We did see that here. John said that sometimes, the discussion questions are too babyish for them, so we highlight every part of the data that are relevant to this code.

The next code is whether the curriculum is culturally or linguistically appropriate or not culturally or linguistically appropriate. And we didn't see that in the data. John did mention that the vocabulary's too hard, but it didn't seem like it was related to the language of the students, their preferred language, so we won't apply that code. The next code is that the vocabulary's too challenging, and we did see that, so we'll code these sentences that say the question vocabulary is sometimes too hard and then the example that explains it. Note here that, you know, these pieces of data, they have multiple codes applied to them because it's relevant for all the sentences in this portion of the interview. And our last code is issues with text and literacy. We heard that at the top that some of the students don't want to read or be read to, and all the lessons involve reading and writing.

So, this shows you how you would go about coding your data and how you would double code if necessary. Our next step happens after you've coded all your data. You might find that you need to group data from related codes together. Like I said, some of the codes might be too specific, and they may be more meaningful when they're grouped together by theme. So, prior to grouping coded data, you're going to want to make sure to review some of the data in each specific code to make sure that they're related to the data in the other codes to check to make sure you should actually group these data by theme. You might need to group or regroup codes multiple times. As you're coding and trying to make sense of your code, you're going to likely find that other codes are interrelated, and they're more meaningful when combined. And you might engage in this process a few times.

I have an example of how to group similar codes that are pretty specific, but all related to each other. The five codes are the curriculum's not appropriate for student's age, the curriculum's not culturally or linguistically appropriate, the vocabulary is too challenging, there are issues with literacy and with text, and there is trouble with some examples in the curriculum. Each of these codes is very specific, but they're similar because they point to an issue that the curriculum is somehow mismatched to the participants. All of these data under the specific codes can be

grouped under a broader code, which is called misalignment of curriculum.

Okay, our sixth step happens after you've coded your data and grouped them thematically. You're going to want to think about what your findings are to meaningfully answer your research questions. So, even after you've grouped your data from more specific codes to broader codes, you might still end up with a really long list of codes, and reporting a list of 15 codes to answer a research question isn't really best practice. It's really difficult for your audience to digest a long list of codes, and you're putting the onus on your audience to figure out what's the most important or meaningful from that list. So it's our job to do the heavy lift for our audience and define what is important and what is meaningful. To develop key findings, you have to reflect on the data and your research questions, and you might consider doing a combination of different factors to decide what to elevate as a finding.

You could take a numeric approach and examine the extent to which codes affect a large number of facilitators, sites, or participants. You could scan and see how often codes appear across your data set, and it's really up to you as the evaluator to determine what this large number is. So, in some cases, it might be 20 percent of participants. In others, it could be three-quarters. In addition to looking at the frequency, you might consider if there are codes that are particularly influential or important for a substantive reason regardless of the amount of times it appears. And I promise I will go through some examples that I hope will help clarify some of this. Determining what a finding is is one of the most difficult parts of conducting qualitative research. Unlike with quantitative research, there aren't specific statistics that are best practices to use.

You really want to determine your findings by choosing the codes with the data that best answer your research questions. Like I said, it's up to the evaluators to define how they'll determine what findings to report. I recommend creating some decision rules or a logic that you'll apply to make decisions about what is important and meaningful to report. The way to create decision rules is to reflect on prior literature, your research questions, your coded data, and create kind of tailored logic and rules to decide what's most important. And I know, this sounds a little bit vague, and I think it's because it's really rooted in your specific study. I'll go through an example and walk you through how I determine what is a decision rule to make. But really, whatever decision rules that you make, the key is to systematically apply them and, in your methods section of your report, document how you determine what is considered a finding. You'll create this process for determining what findings are after you finish your coding. So, it's really important to know what's in the data to

know what findings to lift out. It's not good practice to create decision rules in advance.

Keep in mind that determining findings from your coded data requires multiple readings of the data. So, by the time you're at this step, you've probably read your data a lot of times and are really, really familiar with them, but it's still good practice to review your data again with this fresh perspective of trying to determine what a finding is. We have our example here of how to determine what a finding is and how to figure out what codes to lift up as findings. I'm going to warn you in advance that I'm about to share a table that looks pretty messy, but I will walk us through each part of the table. So, here's our messy table with a lot of text. At the top, we have our research question: what are the key challenges to implementing the curriculum?

On the left, we have our codes, and in the middle three columns, we have some counts. So, the first middle column is the number of respondents. So, these are the number of, let's say, facilitators who said that these codes were an issue for them. In the middle, we have the number of cohorts that were affected, and these are the number of cohorts of students that were affected by these challenges. And then, to the right of that, we have the number of individual students who are affected by these codes. When you use your qualitative analysis software, the number of respondents column will be calculated for you by your coding. But the number of cohorts and the number of students, those were things that I calculated on my own. And you might have similar criteria or additional criteria that you want to include and just keep in mind that some of them will require different sources of data to be able to calculate those numbers. So, you might not have access to all of these pieces of information.

On the right, we have our substantive issues, and these are the reasons that, you know, based on literature or based on our experiences, that we think are important or why these codes might be important to lift up. I want to stop and say that I created this table as an illustration for this webinar. It's not necessary for you to create this when you're doing your analysis. Some people might find it useful, but other people might find it kind of time consuming and not super useful for their purposes. I just wanted to make that note. Okay, so, when we want to create our vision rule, we're going to look at all of the pieces of information that we have, and then we're going to try and decide what is important to lift up based on the knowledge that we have about the data and the program that we've implemented. So, when I'm looking at the frequencies, I'm realizing that this is really about implementing the curriculum and the challenges that were present with curriculum implementation.

And, to me, it doesn't matter quite as much the number of respondents who reported it. What matters the most to me are the number of students that were affected, because implementing the curriculum for my purpose is the whole point—that the students receive some sort of benefit from the curriculum. And so, I would focus more on the number of students affected, and, since I want the majority of students to benefit from the curriculum, I would create a decision rule where at least half of the students would need to be affected by a challenge for me to think that it's worth including in the report. In addition to the frequency, I'm also going to say that if there is some substantive reason present, that's also important to report, because that might be something that's informative to other program staff or the field. It might help them in their projects implementing programs that are similar.

So, I'll say that substantive reason is also important if it affected at least one student. So, those are my two decision rules. The first one is that at least half of the students were affected by the challenge or that there is a substantive reason that affected at least one of the students.

Now, I'm going to walk us through, and we can use the decision rules and determine if it is or isn't a key finding. We'll go through each of these. The first one is scheduling and timing. So, it affected five students—that's less than half, because we have 80 students who were served total. And there is a substantive reason. It says, while it's critical to coordinate scheduling and timing to implement the curriculum, scheduling and timing issues did not affect implementation. It sounds to me like the substantive reason is that scheduling and timing issues are important, but, ultimately, it was an annoyance that didn't really hinder implementation. So, it doesn't seem like that's good enough of a reason for me to lift it up in the report based on my decision rules. And so I will not include that as a key finding.

The second code is unresponsive participants. This affected 60 of 80 students, well over half of the students. And then, the substantive reason listed is that participant engagement in lessons, in discussion and activities, is part of the program model. So basically, the developer is saying that participants have to engage in all parts of the lesson in order to receive benefits of the programming. That's a pretty strong substantive reason, and the majority of students were affected. So, I will include this as a key finding in my report.

I'll go to disruptions to sessions. So, 45 of 80 students were affected, so more than half of the students were affected and then, our substantive reason is that disruptions to sessions can impede implementation, so that could cause disruptions for a lot of people in the classes. It could also

prevent people from receiving valuable content. There is a strong substantive reason and more than half of the students were affected, so I will include this as a key finding.

Our next code is late arrivals. None of the students were affected. Already, we know that this won't be included as a key finding because nobody experienced this. But our substantive reason is that developers stated participants must be part of the beginning of each session. So, one of our decision rules is if at least one student was affected by the challenge, and there was a strong substantive reason, we would have included it in as a key finding. But because zero students were affected, we won't include this as a key finding.

With early departures—this is if students leave sessions early—we had 10 of 80 students affected. That's less than half of the students. And in terms of the substantive reason, the program model states that students should be part of all of the content, but students made up what they missed when they left the sessions. So, the developer may have said, students have to be present for the full lesson, but because students made up what they missed, this isn't enough of a reason for me to lift this up as a key finding.

Our last one is curriculum misalignment. So, 30 of 80 students were affected, which is less than half. Our substantive reason is that the curriculum must be aligned to the target population. Basically, maybe the developer is saying that if the curriculum isn't aligned to the target population, there are some key missteps or there are things that they might not understand or, according to OPA's core themes, we have to have curriculum that's culturally and linguistically aligned to the target population. So, if it's not aligned, then, maybe there's some ways that they can't understand or really fully engage with the curriculum, and this is a big substantive reason. So, even though fewer than half of the students were affected, I will include this as a key finding. We see now that we had a list of six codes, and we applied our two decision rules to whittle down to three findings, which I think is a more digestible number for an audience to read about. And it's showing that you used a systematic process to determine what should be reported and what's meaningful.

Our last step is about reporting our findings. When we are reporting our findings, we want to do so in a way that answers our research questions. We want to make sure to clearly state the meaning and the importance of the findings. Like I said, this is what people mean when they say, tell the story of your findings. It's to convey the importance and significance of the findings. A good check is to ask yourself, you know, why does this finding matter? You should be able to respond to that question, what is the significance of this finding? Why is it important? You're going to want to

use language that accurately reflects the message that were used. We're using qualitative methods here, so, we really shouldn't be using causal language. Causal language is language that directly links programming to outcomes. I have an example here of some causal language. We'll see that the programming is directly linked to some participant outcomes.

“The program did not positively affect participant's reduction strategies for risk behavior because the participants struggle to understand the curriculum's difficult vocabulary.” This sentence is making a direct link between participant's difficulty with the vocabulary and their reduction strategies for risky behavior. Our methods don't actually allow us to make claims like this, so we will not be doing that. What we'll do instead is use language that's clear about the methods that we use and that's not making any direct links. We'll say, “Participants reported they did not understand reduction strategies for risk behavior because they struggle to understand the curriculum's difficult vocabulary.” You're saying similar things, but you're not making direct links, and you're noting that this is based on participant reports. When you're reporting findings, in addition to the three things that we've discussed previously, you'll also want to play to the strengths of qualitative data. This will help the audience better understand or connect to the findings. So, some of the strengths could be that you use illustrative quotes or examples.

You can provide details on background information and context, and, if your IRB [institutional review board] allows, you can provide sample writings, drawings, or images if these are applicable to your research questions. All of these extra details and nuance really helps bring some flavor and life to what you're reporting.

So, our last example for today is an example write-up. There's a lot of text here, so I'll read it and then I'll walk us through some of the different aspects of reporting qualitative research that are included in the findings.

“A key challenge the program team faced when implementing the curriculum was the misalignment between aspects of the curriculum and student's needs. One time of curriculum misalignment was that the language was too advanced for students. Two of the five facilitators stated that they spent at least 10 minutes every hour-long session going over complex vocabulary such as brusque and anachronistic that were only tangentially related to the content. The time spent teaching definitions reduced the time for mentorship activities that students reported were the most valuable because they could connect the dense content to their own experiences. As one student reported, the best part is talking in our mentorship group. ‘That's when I'm like, Oh, that kind of convo happened to my friends and me, too.’ Facilitators address this challenge by—”

and then, you would continue to describe the ways that these challenges were addressed. Now, I'm going to walk us through some of the facets of writing up qualitative research. One of them is to use language that reflects the methods used. You can see what's highlighted. We're saying that these data or these findings are from qualitative sources like interviews and focus groups, and we're showing that by saying that two of the five facilitators stated that they experienced something that students reported was the most valuable and by providing that a student reported this quote. By saying stated and reported, you're using a little bit of language that reflects that these were from interviews and focus groups.

Another aspect of reporting qualitative data is to use examples and quotes. Here, we show what the complex vocabulary is, and, by giving these examples, you can see how they're tangentially related to the content. I mean, they're not actually about the content. So, brusque and anachronistic, you don't need those words to teach the programming. These are really secondary vocabulary terms that are maybe more illustrative. And with the quote, you can see what students feel about these group mentorship activities—that it's the best part—and also how they're connecting these mentorship activities or how they're able to make connections about the content to their own experiences by highlighting conversations that they have with their friends. The examples and quotes aren't really necessary to convey what the finding is, but they add to a reader's understanding, and they can kind of help ground your findings in very concrete examples.

The third thing I wanted to highlight about reporting findings was stating the significance of the findings. A critical practice when reporting findings is to show why they're important. In this example, the significance is that the challenge of complex vocabulary interfered with a really important and valuable part of the programming. It's not just that it took 10 minutes out of a 60-minute long session, but that it was taking time away from what students found to be the most valuable, the group mentorship.

And this concludes the presentation portion of the webinar. Now, we can move on to any questions that you have based on what I presented or any of the questions that you sent in advance, I think, Diana, you're going to be reading the questions, is that right?

Diana McCallum:

That's right, and we do have a couple coming in. So, we'll transition to those. Just so that everybody knows, you can submit your questions by going to the Q&A widget that's at the bottom of the screen, so feel free to submit your questions and we'll go through them.

I did want to make one point for folks that may be wondering how some of the strategies that Jane talked about may apply in the context of continuous quality improvement. I just want to be clear that you can use these processes that Jane talked about in the context of your CQI efforts, and they're one strategy for helping you do your CQI systematically. But you don't have to use these strategies for CQI, and there are certainly other ways that you should feel free to add on the data that you're gathering for CQI and assessing what changes you might need to your program in other ways. But we do want to be clear that the strategies that Jane has talked about are definitely ones that you should consider when you're doing a research project, and they're particularly helpful as you're beginning to analyze your data and preparing to craft your journal articles or start your qualitative. I just wanted to be sure to make that point.

We have a couple of questions. Jane, the first one, someone has asked, for planning purposes, do you have hard and fast rules for how long it takes to code qualitative?

Jane Choi: Sure. So, I would say that coding goes pretty quickly if you have a really strong coding structure in place and if you have clean data. A lot of the qualitative data you collect will probably be from interviews or focus groups. And those conversations happen in real-time, which happens slower than how fast we read. And so, coding usually will be about half the speed I would say. If you have an hour-long interview, it might take you about 30 minutes to code it. But this is kind of on average across your interviews. The first few interviews you code you'll find to be pretty slow, and it really speeds up after that.

Diana McCallum: Alright, so, there is another question here. Is it better to develop codes before reviewing the data and the person says if the answer is yes and if so, this is an area where quantitative and qualitative data analysis may differ, right? But Jane, can you talk a little bit about that?

Jane Choi: Sure. I would say you want to develop codes before reviewing the data, after you collect data, and while you're coding. So, usually, you'll be doing both. In practice, usually evaluators will develop codes after they've collected their data, so they already have some idea of what's in the data, even if they're creating codes beforehand. I would say that there isn't really a better or worse, but, usually, researchers will do both.

Diana McCallum: Great, thanks, Jane. And I just want to encourage folks to continue to submit questions within the Q&A widget while we're here in the webinar. There is one other question that kind of dovetails with the point I was making about CQI and the question is for Jane, whether you have any recommendations for analyzing qualitative data for CQI purpose, you

know, as opposed to publishing, do you have any strategies or any suggestions about that? That'd be helpful.

Jane Choi:

Yeah. I'm going to echo what you said, Diana, that I think when you're doing CQI, you're trying to use your data really rapidly to make decisions midstream while you're implementing the program. And so, I think that what I described today probably takes too long to be very useful for a CQI process. If you have any high priority issues that come up as you're collecting your data, you should use your formal CQI processes that are in place to act on those high priority issues. If you had more flexibility to do the type of analysis that I described today, one of the things that could make it go faster is randomly choosing 20 percent or a subset of the data to do this type of analysis and then analyze the rest later for publishing. But everything Diana said earlier is right: if you have CQI processes in place, you shouldn't abandon those to use this analytic method because there's a difference in needing to respond midstream to changes to improve programming.

Diana McCallum:

Great, thanks, Jane. Here's another question, and this is a larger one. It's important, but, the person is saying, it seems hard to have qualitative results that are unbiased and objective. Jane, it'd be helpful if you could speak to how people make sure that their results are trustworthy. Do you have any suggestions about approaching that?

Jane Choi:

Yes, I do. I think the best way to make sure your results are trustworthy is to have the systematic processes to guide how you analyze your data. This includes focusing all your analysis on your research questions, having a strong and clear coding structure in place, training your coders if you have multiple coders, and conducting regular checks to make sure everything's going consistently. And, of course, having a clear logic or decision rules to define how you'll lift up codes for your findings. I think reporting your methodology in your report is a way to be kind of transparent with your audience. But I also want to speak to the point on, you know, it being difficult to have qualitative results that are unbiased and objective. I think the processes that I mentioned will help limit some of the bias and kind of make sure that you're being systematic.

But with qualitative analysis, you've seen that there's a lot of decision making involved and a lot of researcher input, and there's just so many human elements involved in every step of the process from data collection to the analysis. So, I think the goal here is really to be systematic and transparent about all the decisions you make and the steps that you took and clearly reporting this in your report. You don't want to hide anything.

You want to show where your fingerprints are and make that really apparent so that your reader can make a decision about whether they should trust these results. I think the things that I read that make me a little bit skeptical are when the methods section is very vague and opaque, because then I'm not really sure what steps or what processes people took. And a lot of what's hidden could have affected how you derived your results.

Diana McCallum: Thanks. Please continue. We certainly have more time for questions, and then there's a few other have rolled in here.

Jane Choi: We did have a good suggestion from Nicole about how to use qualitative data for CQI purposes, and she says it's a great time to use post-it notes, stickies, and dot voting. I agree. Those are ways to involve everybody and go through some of your data quickly.

Diana McCallum: Great. There's another question: if I'm including quotes or examples from respondents, how do I protect their confidentiality and reporting on those quotes?

Jane Choi: Yes, that's a great question. A lot of times, you'll have some confidential information or personally identifiable information that gets slipped into qualitative data. And a lot of times, your participants will drop names of locations, sites, or even people. So, if you're using quotes or examples, you want to make sure to very carefully comb through them and remove confidential information. You could apply pseudonyms to sites or to participant names, or you could provide descriptions of things if it's important to understand the content. For example, if they're describing the context and location is important for that, instead of saying the city's name, you could say a description of it, like this is a small, rural Midwestern town. I do want to say that, some IRBs are really strict. When I used to be part of a university setting, I had an IRB that required that I get explicit permission to use quotes. You want to make sure that your IRB is alright with you using quotes and some specific details.

Diana McCallum: Great, another question. I feel like this is pretty common, but how do you decide between qualitative versus quantitative data collection? When is using a survey a better option?

Jane Choi: Yes, that's a good question. I think it really depends on your research question. Usually, for program evaluations, we'll use quantitative data to understand the outcomes—the impact or an effect of a program on specific measurable outcomes. And, in that case, you'll probably want to use a

survey or administrative data. Qualitative research, on the other hand, provides a lot of details and nuanced information. For example, if you have an impact study, and you're measuring your outcomes quantitatively to see if a program had effects, you might not understand why or why a program did or did not have effects. Qualitative data can help provide some of that information. You can document experiences of staff and participants and contextual factors and better understand what happened with program implementation to help explain why there were or why there weren't effects. I think the other benefit of qualitative research, which I described in the reporting stage, is that some of the data elements like quotes and details can help make your reports more accessible to readers and sometimes even more interesting, because you're able to share some of the stories or some of the experiences of your respondents.

Diana McCallum: That's great, thanks, Jane. And I think this next question, this is another one we hear a lot about sort of how intensive it is to train coders. This person is asking whether you have any recommendations for training coders.

Jane Choi: I do have recommendations. I would say to hold a pretty strong initial training where you're going through your research questions, any software tools you're using, and each of the codes, and I would go into pretty good detail about what each code means and when to apply it and give some examples. And then you would want to have your coders code one of the exact same interview together and then, maybe individually, code it and then, check how everybody did on their coding to make sure that everyone's applying the codes consistently. And then you're going to want to keep going through that exercise a couple times through the coding process. After everybody codes 10 or so interviews, you might want to revisit and make sure everyone's still consistency applying the codes. I think coding can sometimes be a little bit boring and labor intensive, so all of the good things you can do to keep people engaged, like have food at the trainings and somehow make the trainings fun and engaging for everybody, are beneficial. I think those are all like good practices to keep people motivated.

Diana McCallum: That's helpful, thanks, Jane. So, that about wraps up the questions we have so far. If you have other questions, feel free to submit them in the Q&A widget. We're happy to continue. But Jane, I think they want to transition to the resources you were going to highlight for folks. This may be a good time to do that and then we'll see if any other questions come in.

Jane Choi: Alright, that sounds good. I have some resources here. The first chunk are readings on qualitative analysis, and we have some resources that are like textbooks. The first one, I actually think there's a much more updated version of this now. I think there's a 2018 version, so, I can update my slides. And then we have some sources from the federal government from NIH, and then the last resources are from Health and Human Services. And this resource is actually intended for federal staff such as program officers, but I think there's a lot of good background information of when to use qualitative research and what are some of the theoretical aspects of qualitative research, and it's written in really practical terms. I think this one is also a great resource for background reading. I have some examples of qualitative analysis programs. I've used all of these. They vary in price and in what they can do. But they're all pretty good, and I wanted to give some examples in case you're not familiar with any of the qualitative analysis programs. And there are plenty more out there. So, if you have one that you prefer using, by all means, you should continue to use that one. And that's pretty much it for the resources.

Diana McCallum: I think there is one additional question. I know, it was close to something we've covered, but what's the right level of detail? He's asking, what's the right level of detail in developing codes?

Jane Choi: Okay, good question, I think this is a little bit based on personal preference of how you think and how you want to keep your codes and your data organized. My personal preference is to go extremely specific. For me, I find it faster to assign really specific codes, and it's easier for me. I don't really have to think as hard if I'm just assigning really specific codes. So, what I'll do is assign specific codes, and then later I'll group them into broader codes. I do know people who prefer to go the opposite direction; they'll group everything into a broad code—they might group all of the relevant pieces of the interview into a code that says, "Program challenges." And then they'll open up all the data in that bigger code and apply really specific codes, like issues with curriculum, issues with scheduling, issues with attendance. So, I think a little bit of it is personal preference, but I found that it's easier and faster for me to go specific and then the group rather than to go broad and disentangle them later.

Diana McCallum: Great, thanks, Jane. I think that's all we have for questions. So, as folks continue to think of other questions, you can certainly email the TPP Evaluation TA Help Desk. The email address is TPPEvalTA@Mathematica-mpr.com. If you have other questions, you can certainly reach out to us. I know, there are a few grantees on the line that work directly with Jane and the other TPP team liaisons, so feel free to

send questions about your qualitative analysis directly to them if you have future calls, but we certainly welcome any other questions around this topic. And it doesn't look like there are any other questions, so I think we can go ahead and close out the webinar. Thanks, everybody, for your attention today and for the great questions. Thank you, Jane.

Jane Choi:

Great, thank you so much.