9-24-2018

When Interviewing: How Many is Enough?

William W. Cobern
Western Michigan University, bill.cobern@wmich.edu

When interviewing: how many is enough?

William W. Cobern
Research Seminar Paper
The Mallinson Institute for Science Education
September 24, 2018

Researchers need to know what is an acceptable *number* for interview work

How does one decide the acceptable number of people to interview? Or one might ask what is an acceptable N size for interview work?

I raise this issue because the N size, or interview numbers, was questioned by a couple of people at a student's dissertation proposal presentation. The student proposed to interview 20 parents of primary, middle, and high school students. The proposed 20 parents thus being divided over three grade bands. This proposal was questioned by at least two people as having too few parents in each band. I don't recall if the word "generalization" was used but the gist of the objections was that dividing 20 across three grade bands would mean too few subjects in each band for generalization. Subsequently the student and his committee decided he should focus on only one grade band but not for reasons having to do with generalization.

With respect to interview work, the concept of generalization is misapplied. Thus, on this point the student's objectors were wrong. As it happens, the concept of generalization is often misapplied to qualitative research and probably more so by people who are primarily quantitative researchers. Still, there is an underlying question: what is an acceptable number for interview work?

What follows is a brief explanation of why the concept of generalization is inacceptable in qualitative interviewing, what wording to use in place of generalization, and how one should decide on interview number.

1) The related concepts of generalization and N size are from quantitative work. They have no counterparts in qualitative research including qualitative interviewing. Generalization is a statistical concept that is typically expressed as a p-value. In other words, generalization is a probability function; the probability that a null hypothesis is true. One often sets p-value at <=0.05 for statistical significance. What this 0.05 means is that if you were to run *exactly* the same test situation 100 times, and the null hypothesis is true, then you should get the same result or a more extreme result only five times, which suggests that the null hypothesis is probably not true. You have *generalized* from one test to 100. If you are testing an instructional innovation for statistical difference with respect to a control, another way to look at what p-value means for p<=0.05 is that if 100 teachers implemented your innovation (and they did it exactly the way you did), you would expect that only about five teachers would not get a better result than they would have using the control instructional approach.[1] You expect the others to get better results. You have generalized to 100 teachers.

2) N size is related to the statistical concept of generalization through power calculations. Admittedly, researchers often neglect this calculation (typically because they are using *convenience* samples), but power calculations are used for estimating the N size needed to show statistically significant difference if such a difference exists. Here is a quote from a web page about statistical procedures.

> "In plain English, statistical power is the likelihood that a study will detect an effect when there is an effect there to be detected. If statistical power is high, the probability of making a Type II error, or concluding there is no effect when, in fact, there is one, goes down… Statistical power is affected chiefly by the size of the effect and the size of the sample used to detect it. Bigger effects are easier to detect than smaller

effects, while large samples offer greater test sensitivity than small samples."[2]

As you can see, the ability to detect a true effect sensitive to sample size. Hence, the ability to generalize is sensitive to sample size. But in qualitative work such calculations do not exist and therefore the concept of generalization cannot be applied to qualitative work.

3) In qualitative work rather than speaking about generalization one should use forms of the word "indicative" or similar words such as "suggest." You would say something like "the findings of this study are *indicative* of what one would find in other situations given similar characteristics." Or, "this study indicates that…" Or, "this study *suggests* that…

Furthermore, don't use the word generalization when you talk about the limitations of your qualitative work; use the language of indication or suggestion. The limitation is that your findings are indicative only for situations having similar characteristics.

4) But I still haven't answered the question of how many to interview, and the number does matter though not for the same reason number matters in quantitative work. In a quantitative study let's say you use a Likert item survey. For this quantitative, Likert item study you need numbers so that you can estimate how likely it is that people (of similar characteristics) will hold the opinions represented by the items.

In contrast, an interview is used to determine *what* the opinions are. Hence, you need to interview enough people so that you learn most if not all possible opinions (amongst people of similar characteristics). Yes of course you may want to know which opinions are more popular but that's not the primary aim of qualitative work.

For qualitative interviewing there is a critical assumption: the number of unique opinions is not very large. Yes I know that if you ask two MISE professors a question you will get at least three opinions. LOL

But seriously, if we asked professors what they thought about working at WMU there would be a limited number of opinions. If you ask 100 professors you are not going to get 100 unique opinions. What you will find is that several opinions get repeated over and over. In my student's dissertation proposal, when he speaks with parents about NGSS he will encounter a limited number of unique opinions. What he will find is that as he goes down his list of parents, a few opinions will keep reoccurring. Opinions on a subject don't run in the hundreds; they don't even run in the dozens. On any given topic there are typically a handful of unique opinions. There just are not that many opinions that one could hold about most topics. The goal of qualitative interviewing is to capture most if not all opinions however many opinions there are. And this is where the number of people needed for interviewing comes into question.

5) Clearly, the likelihood of capturing most if not all opinions increases with the number of people one interviews. The thing is, once you have captured the possible opinions, there is little reason to continue interviewing more people. You have reached "saturation." Interviewing more people will not result in more opinions because very likely there are no more opinions. But how many?

One approach to how many, and it is one that I've used, is that you don't estimate ahead of time how many people to interview. You keep interviewing until you reach a point where you stop getting unique opinions and all you're hearing is what you have heard from previous interviewees. At that point you interview maybe one, two, or three more for insurance, but you have reached the number you need. In a study I published quite a few years ago, I quit at 16 interviews.

On the other hand, often times for logistical reasons, time constraints, and financial ability to pay honoraria we have to decide ahead of time the maximum number of people to interview. This of course is the situation in which most student researchers typically find themselves. So if you think about the opinions that people hold on any topic you might think there just two opinions… Or, are

there three to five opinions? Could there be 10 distinct opinions on this topic? The literature can help you because it can suggest what opinions might be out there but conventional wisdom (maybe we would even say common sense) is that on most topics there are not 10 unique opinions.

So if we assume that there will be no more than 10 unique opinions on most of the topics we would want to ask people about, how many people do we need to interview to get those 10 opinions? That is the question you have to answer. That is, that you have to estimate an answer for that question. That estimation gives you the number of people you should plan to interview. Conventional wisdom suggests that that number is between 15 and 20. It's a good bet that the high school students' parents will have fewer than 20 unique opinions about NGSS.

For the student (or any researcher), it is not the number of people one plans to interview that is the question needing to be answered. The important question is how likely are parents of students in the three grade bands to have different opinions so that the domain of unique opinions across the three grade bands exceeds the number of unique opinions in anyone grade band. If it can be argued that grade band is unimportant, then his original plan was fine. On the other hand, if the grade bands are likely to result in different opinions, then six or seven interviews per grade band may not be enough to reach saturation per grade band. Probably too risky.[3]

6) Closing thoughts.

A. When you report the results from qualitative interviews you should report whether or not you reached the point where you were hearing the same things from interviewees.

B. Whether you are doing qualitative or quantitative work it is extremely important that you adequately describe the characteristics of your study situation. For example, if you're going to interview parents we need demographic information about the parents and about the schools that their children attend. Generalization in quantitative research means generalizing to *similar* situations. In qualitative work, the results are indicative of what one would find in *similar* situations.

I'm sure I have oversimplified things, but perhaps what I've written will still be helpful.

[1] If the wording of this section seems awkward it is because statistical testing is about the probability of a null hypothesis being true. Low p-values suggest that the null hypothesis is not true, but this does not necessarily mean that the treatment hypothesis is true.

[2] https://effectsizefaq.com/2010/05/31/what-is-statistical-power/

[3] Here is some example wording suggested by my colleague Dr. Brandy Pleasants:

Based on research with a similarly homogenous group it seems that about 10 participants is sufficient to cover all reasonable responses I might get. I therefore plan to interview no less than 10 participants, with a goal of 15 (even if saturation is reached) if I am still seeing variation in the data I will continue to interview until I reach saturation.

Suggested citation:

Cobern, W. W. (2018). Notes on what is an appropriate number for interview work? Kalamazoo, MI, The Mallinson Institute for Science Education, Western Michigan University: 1-3. Retrieved from https://scholarworks.wmich.edu/science-seminars/1.