**SATURATION**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

In qualitative research, researchers are not always sure of the amount of data they need to collect. Should they, for example,

* interview 10 participants or 30 participants?
* collect 5 hours or 50 hours of interview material?
* observe a community for 10 hours or 40 hours?

To decide the amount of data to collect, many qualitative researchers utilize the principle of saturation. According to this principle, researchers should continue to collect data until the participants no longer impart additional information—information beyond the insights the researchers have already acquired. Although many researchers advocate the benefits of saturation, scholars have not reached consensus on how they should decide whether saturation has been achieved. This document offers some insights to help researchers reach this key decision.

|  |
| --- |
| **How much data should researchers collect?** |

**Previous recommendations**

Scholars have not reached consensus on the amount of data that researchers should collect. The following table, for example, presents some recommendations from various scholars.

|  |  |
| --- | --- |
| Author | Recommendation |
| Bertaux (1981) | * Qualitative studies should entail more than 15 interviews |
| Creswell (1998) | * Studies in phenomenology should comprise 5 to 25 interviews |
| Morse (1995) | * Studies in phenomenology should comprise 5 to 25 interviews * Studies in ethnography should comprise 30 to 50 interviews * Grounded theory should comprise 30 to 50 interviews |
| Mason (2010) | * Qualitative PhDs tend to include 1 to 95 interviews |

Even this brief table implies that

* the recommended number of interviews—or amount of data—that needs to be collected in qualitative research varies widely across scholars
* the amount of data that needs to be collected probably depends on many considerations, including the methodology that researchers adopt and the responses of participants (Baker and Edwards, 2012)

**Results of past studies**

Some researchers have undertaken empirical studies to assess the number of interviews that researchers should conduct. To illustrate

* as Morgan et al (2002) showed, over 80% of the themes can be extracted from the first 10 interviews. Almost all the themes can be extracted from the first 20 interviews
* if at least 55% of participants adopt some belief or theme, six interviews is enough to be 99% certain of identifying this belief or theme (Galvin, 2015)

**Importance of sample size**

In quantitative studies, when the number of participants is inadequate, several problems can unfold

* First, the researcher is unlikely to generate significant results—that is, p values less than 0.05.
* Second, the results might not be reliable. That is, if the study was repeated in the future, the results might be quite different.
* In both instances, researchers cannot reach firm conclusions about the data.

In qualitative studies, however, the implications of inadequate participants or insufficient data are different. So, what consequences unfold if researchers do not interview enough people or collect enough data? Is this topic important? According to Onwuegbuzie and Leach (2007)

* if the data is inadequate, the researchers may overlook important nuances—such as variations and properties of the themes or categories as well as insights on how these variations and properties vary across contexts
* without this knowledge, readers cannot decipher the extent to which these results or conclusions are likely to apply to other contexts—that is, the results are not readily transferable

|  |
| --- |
| **Is saturation the solution?** |

Rather than specify the number of participants they will interview in advance, researchers often decide they will continue to recruit participants, and to collect data, until they reach saturation. Roughly speaking, saturation is the moment in which participants are no longer imparting additional information—information beyond the insights the researchers have already acquired. That is, in essence, the participants are repeating the responses of previous respondents.

**Saturation from the perspective of grounded theory**

The notion of saturation emanated from grounded theory. According to proponents of this this theory

* researchers strive to develop a theory of some phenomenon—a theory that specifies how various themes or concepts, called categories, are associated with each other
* the researcher continually updates decisions on which data to collect, which participants to interview, and which questions to ask.
* as some moment, they feel the properties of these categories—and the associations between these categories—are no longer shifting as they collect more data. That is, the data confirms, but not does extend, the theories they have developed thus far.
* this moment is called saturation
* because the theory is not shifting, this moment is sometimes called theoretical saturation.

Proponents of other philosophies and methodologies have borrowed this notion of saturation. These researchers, however, tend to adjust the definition of saturation to suit their methodologies. For example, researchers who espouse thematic analysis might define saturation as the moment in which the themes they are generating are no longer changing as they collect more data.

**Limitations of saturation**

Nevertheless, because of several complications, researchers often challenge the merits of saturation. The following table outlines these complications.

|  |
| --- |
| Complications of saturation |
| **Metaphor exaggerates progress.** The word saturation implies that researchers have reached complete insight about some topic or phenomenon. Yet, in principle, further insight is always possible.  Perhaps a more suitable term is *sufficient depth* (Dey, 1999) or *conceptual depth* (Nelson, 2017). That is, in practice, the researchers should attempt to develop the categories or themes to a sufficient extent to be confident their conclusions are likely to be transferrable to other similar circumstances |
| **Ambiguity outside grounded theory**. Because the principle of saturation was developed with reference to grounded theory, the precise meaning of saturation to proponents of other methodologies is uncertain |
| **Inflated quality.** To show their results are transferable—and thus to demonstrate the quality of their work—researchers like to maintain they have reached saturation. Yet, this claim is typically flimsy, because no definitive criteria has been developed to measure saturation. |

|  |
| --- |
| **How to demonstrate saturation 1: The comparative method for themes saturation** |

As the previous section implies, if researchers could measure and monitor saturation more systematically, the principle of saturation might be more compelling. Constantinou et al. (2017) developed a simple, but methodical, set of procedures to gauge saturation. The following table outlines this procedure. This procedure overlaps with the activities that most qualitative researchers conduct anyway but formalises choices around saturation.

|  |  |
| --- | --- |
| Activity | Details and illustrations |
| **The first interview** |  |
| After the first interview, code every segment of data | * Most qualitative researchers like to assign each segment of data—such as each distinct sentence or proposition—a code * Codes are often 1 to 5 words * A typical interview might generate 20 to 50 codes * For example, if a researcher wanted to explore how the confidence of research candidates changes during their candidature, typical codes might be “doubted abilities”, “questioned assumptions”, and “felt special” |
| After the first interview, arrange the codes into overlapping topics | * Researchers often develop a codebook—a list of all the codes they generated * They could then arrange these codes into clusters * For example, one cluster could include all codes in which research candidates described unpleasant thoughts about themselves |
| After the first interview, derive themes from these clusters of codes | * Most qualitative researchers will generate themes—often abstract labels—that describe the shared features or essence of overlapping codes. * For example, codes in which research candidates experienced unpleasant thoughts about themselves could be reduced to one theme, such as “personal doubts and uncertainties”, or perhaps more than one theme * This procedure might generate between 10 and 20 themes, for example * Depending on the methodology, you might refer to these themes as categories, concepts, and other labels |
| **The second interview** |  |
| Then apply this procedure—that is, generate codes and derive themes—to the second interview |  |
| Count the number of shared themes and unique themes | * To illustrate, the first and second interview might generate 15 shared themes * The second interview might uncover 2 themes the first interview did not identify, called new themes |
| **Later interviews and saturation** |  |
| After each interview, continue to apply this procedure |  |
| If approximately three consecutive interviews generate no new themes, you have probably reached saturation—sometimes called a saturation threshold |  |
| You can refrain from further interviews |  |

**Limitations of this comparative method for themes saturation**

Although simple, this method does present a few complications. The following table outlines these complications

|  |  |
| --- | --- |
| Limitations of this method | Details |
| The precise saturation threshold is ambiguous | * To illustrate, researchers could conclude they have reached saturation after two, three, or more consecutive interviews generate no new themes. * But should this number be two, three, four, or what? |
| The order in which researchers code interviews significantly affects the saturation threshold | * For example, if researchers arrange participants chronologically—and thus code the first participant, then the second participant, and so forth—they might reach saturation after 5 participants. * If they code participants alphabetically—such as code Adam, then Burt, and then Carla, for example—they might reach saturation after 10 participants and so forth. |
| This approach does not apply to grounded theory | * To illustrate, the number of themes or categories might plateau after 10 or so interviews * But the properties of these categories—and relationships between these categories—might continue to evolve after even 20 interviews * Hence, if researchers applied this comparative method for themes saturation, they might prematurely stop interviewing after 10 or so participants * In short, this method generates theme saturation but not theoretical saturation |

|  |
| --- |
| **How to demonstrate saturation 2: A method that utilises information thresholds** |

Guest, Namey, and Chen (2020) developed an approach that is similar to the comparative method—except that researchers can adjust the parameters. For example, the researcher can decide the data have reached saturation once the percentage of new themes is lower than 5%, 2%, or 0%. The following table illustrates how researchers can apply this technique.

|  |  |
| --- | --- |
| Activity | Details and illustrations |
| **Examine the first four interviews** |  |
| List the themes you extracted from the first four interviews | * For example, you might subject these data to thematic analysis—a technique that extracts themes from the transcript * You could choose another number, such as five * This number is called the base rate * Guest, Namey, and Chen (2020) showed a base rate of four is usually adequate |
| Count the number of unique themes in these four interviews | * To illustrate, this thematic analysis might have generated **40** distinct themes |
| **Examine the next two interviews** |  |
| List the themes you distilled from the next two interviews—but only themes you had not distilled before | * You could choose another number, such as three * This number is called the run length * Guest, Namey, and Chen (2020) showed a run length of two or three is usually adequate |
| Count the number of unique themes in these two interviews | * To illustrate, the thematic analysis might have generated **5** new themes |
| **Examine the saturation ratio** |  |
| Divide the number of new themes in the last set of two interviews by the number of themes in the first four interviews | * = 5 / 40 * = .125 * This value is called the saturation ratio |
| Decide whether this saturation ratio exceeds 0.05 | * If the saturation ratio exceeds 0.05, conclude the number of new themes in the last two interviews was high—and hence saturation was not reached * If the saturation ratio is less than 0.05, conclude the number of new themes in the last two interviews was low—and hence saturation was reached * Researchers can use a lower number than 0.05—and may even use 0—as the threshold |
| **Repeat with additional runs** |  |
| If saturation was not reached, repeat this procedure with two more interviews | * For example, suppose only 1 new theme emerged from the seventh and eighth interview * The saturation ratio = 1 / 40 = .025 * This value is less than .05, indicating that saturation was reached * Because the last two interviews did not produce more than 5% of the themes, the researcher will conclude that saturation was reached after the sixth interview |

**How to report these results**

The researcher then needs to report these results. Guest, Namey, and Chen (2020) devised a format that researchers could follow. Specifically, the researchers could write something like the following box. In particular

* 5% is equivalent to 0.05
* six refers to the number of interviews before the last run
* +2 refers to the run length

|  |
| --- |
| When the base size was 4, the researchers reached the 5% new information threshold at 6+2 interviews. |

|  |
| --- |
| **How to demonstrate saturation 3: Conceptual depth criteria** |

To address the limitations of other methods—especially if you plan to conduct grounded theory—you can apply the notion of conceptual depth criteria, developed by Nelson (2017). In particular, Nelson (2017) argues that

* researchers who apply grounded theory, or conduct qualitative research in general, can never reach saturation. Further data could always extend their themes or theories
* but researchers can achieve conceptual depth—enough support of their themes or theories to be confident their conclusions can be transferrable to particular circumstances

To reach this conceptual depth, researchers should asses five criteria. The following table defines these criteria and then illustrates how researchers can assess these criteria.

|  |  |
| --- | --- |
| Criteria to achieve conceptual depth | Notes on how to assess these criteria |
| **Range**: Multiple quotes or examples support each category or theme | * To illustrate, several quotes that are almost identical would not alone fulfill this criterion of range—because similar quotes do not clarify the range or scope of some category or theme * The criterion of ranges is more likely to be fulfilled if diverse sources of data—disparate quotes or a blend of quotes and observations—corroborate the categories or themes.   Illustration when a category is a variable   * Sometimes, a category might represent variations across participants, such as “extent to which candidates prioritize learning over status”. * To achieve the criterion of range, the researcher would need to collect quotes or evidence of these variations. * That is, some quotes might corresponds to candidates who prioritize learning; other quotes might correspond to candidates who prioritize status |
| **Complexity**: The various categories or themes in the data should be related or connected to each other in some sense | * The researcher might create a diagram or matrix to show how the various concepts, categories, or themes are related to each other. * If this criterion is not fulfilled, and the categories or themes seem quite isolated from each other, the researcher has probably overlooked some vital insights |
| **Subtlety:** The researcher should demonstrate how the precise meaning of particular words, phrases, or codes varies across contexts | * The researcher, for example, might notice that many participants utilize a specific phrase or broach a specific concept. Research candidates, for example, might often refer to burnout. * But, the researcher should then apply the comparative method to consider how the meaning of this phrase or concept differs across circumstances * Individuals might refer to burnout as a justification to conserve energy, as a concern about their health, and so forth. * Memos about these codes can help researchers unearth these subtle variations in some code or concept across circumstances * If this criterion is not fulfilled, the researcher may have overlooked vital insights and nuances about this category or theme |
| **Resonance**: The themes or theories that researchers generate should be compatible with past literature | * The themes and theories the researchers propose, although fresh and novel in one sense, should overlap closely with existing accounts |
| **Validity**: The themes or theories that researchers generate should also apply to other settings—analogous to external validity or transferability | * If the study revolves around a particular setting, the people in a similar setting should largely agree with the results and conclusions of this study |

In short, according to this perspective, researchers should assess the extent to which

* their categories, themes, or theories are supported by a variety of data sources, called range
* their categories or themes are embedded in a cohesive framework, called complexity
* they have uncovered subtle variations in the meaning of specific words, phrases, or concepts, called subtlety
* their themes or theories are consistent with previous literature, called resonance
* their themes or theories apply to other similar circumstances or settings called validity

Researchers who feel their work does not fulfil these criteria should collect and analyse more data. Researchers who feel their work does fulfil these criteria have reached conceptual depth, comparable to saturation. Unfortunately however, in practice, researchers can assess these criteria only subjectively rather than precisely and objectively.

|  |
| --- |
| **References** |

Baker, S. E., & Edwards, R. (2012). How many interviews is enough? National Centre for Research

Methods, Economic and Social Research Council (ESRC).

Bertaux, D. (1981). From the life history approach to the transformation of sociological practice. In

D. Bertaux (Ed.) Biography and society: The life history approach in the social sciences (pp. 29-45)

London: SAGE.

Constantinou, C. S., Georgiou, M., & Perdikogianni, M. (2017). A comparative method for themes saturation (CoMeTS) in qualitative interviews. Qualitative Research, 17(5) 571-588. Doi: 10.1177/1468794116686650

Creswell, J. (1998). Qualitative inquiry and research design: Choosing among five traditions.

Thousand Oaks, CA: SAGE.

Dey, I. (1999) Grounding grounded theory. San Francisco, CA: Academic Press.

Galvin, R. (2015). How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge?. Journal of Building Engineering, 1, 2-12.

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. PLoS One, 15(5)

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. Forum:

Qualitative Social Research, 11(3): Art.8.

Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. J. (2002). Risk communication: A mental models approach. Cambridge University Press.

Morse, J. (1995). The significance of saturation. Qualitative Health Research, 5(2), 147–149.

Morse, J. (2015). “Data were saturated…”Qualitative Health Research, 25(5), 587–588.

Nelson, J. (2017). Using conceptual depth criteria: Addressing the challenge of reaching saturation in qualitative research. Qualitative Research, 17(5), 554-570. doi: 10.1177/1468794116679873

Onwuegbuzie, A., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in

social science research. The Qualitative Report, 12(2), 281–316.