Article

Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research

Qualitative Research
2017, Vol. 17(5) 554–570

DOI: 10.1177/1468794116679873
journals.sagepub.com/home/qrj

James Nelson
School of Social Sciences, Education and Social Work, Queen's University Belfast, UK

Abstract
Saturation remains a problematic concept within the field of qualitative research, particularly with regard to issues of definition and process. This article sets out some of the common problems with saturation and, with reference to one research study, assesses the value of adopting a range of 'conceptual depth criteria' to address problems of definition and process when seeking to establish saturation within a grounded theory approach. It is suggested that the criteria can act as a test to measure the progress of the theoretical sampling and thus ascertain the readiness of the research for the final analytical stages and theory building. Moreover, the application of 'conceptual depth criteria' provides the researcher with an evaluative framework and a tool for producing a structured evidence base to substantiate choices made during the theoretical sampling process.

Keywords
conceptual depth, education, grounded theory, qualitative research, saturation

Introduction

The concept of saturation can be problematic for those researchers involved in qualitative research who use it as means to determine sample size. O'Reilly and Parker (2012) suggest that the term has somehow slipped into the lexicon of researchers in a way that assumes a shared understanding of the concept, when in reality neither its meaning nor its application could be said to have secured broad consensus among the research community. For several decades now, researchers have drawn attention to the lack of guidance around applying saturation (Guest et al., 2006) and the absence of any 'tests of adequacy' for estimating sample size in qualitative work (Morse, 1995). Mason (2010)

Corresponding author:
James Nelson, School of Social Sciences, Education and Social Work, Queen's University Belfast, 69-71 University St, Belfast BT7 1HL, UK.
Email: j.nelson@qub.ac.uk

has even asserted that many researchers have seemingly misled their readers about the saturation process. Yet, despite the fact that 'it remains nebulous' and the process 'lacks systematization' (Bowen, 2008: 139) it is a common requirement of funders and reviewers of qualitative research (Kerr et al., 2010; O'Reilly and Parker, 2012).

This article attempts to address the issues of meaning and process in relation to saturation in two ways. Firstly, it begins with an analysis of the concept of saturation, its definition, attendant problems and some current responses to these problems. Secondly, having established the issues around saturation, it draws upon a qualitative research study to explore how one researcher sought to address these issues. This includes some description of the context within which the research was conducted and an outline of the actions taken, in particular the development of 'conceptual depth criteria'. Explanations of the 'conceptual depth criteria' are provided along with examples of how they were applied in the research study. In conclusion there is some reflection upon the potential for the criteria to address the problems identified.

Defining saturation

Saturation has its origin in the theoretical sampling process which is part of the grounded theory method of qualitative research. Glaser and Strauss (1967: 45) define theoretical sampling as 'the process of data collection for generating theory whereby the analyst jointly collects, codes and analyzes his data and then decides what data to collect next and where to find them, in order to develop his theory as it emerges.' In practice then, the researcher does not follow a pre-determined path in choosing who or where to collect data from, rather the research moves from research questions to limited data collection to data analysis. Only when some analysis has occurred is the researcher in a position to select the next sample for data collection, and this allows ideas emerging from the analysis to be tested.

In the words of Glaser and Strauss (1967: 61) "'Saturation' means that no additional data are being found whereby the sociologist can develop properties of the category. As he sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated."

Also working within a grounded theory approach, Corbin and Strauss (2008: 263) stay close to this definition when they describe saturation as: 'The point in analysis when all categories are well developed in terms of properties, dimensions and variations. Further data gathering and analysis add little new to the conceptualization, though variations can always be discovered.'

In both cases the emphasis is upon the saturation of 'categories' though Corbin and Strauss in using the word 'conceptualization' indicate that it is conceptual categories that they have in mind. Going further, Charmaz (2014: 213ff), provides her readers with a more detailed definition as part of a broader discussion of saturation. She defines saturation as the point at which 'your categories are robust because you have found no new properties of these categories and your established properties account for patterns in your data… you have defined, checked, and explained relationships between categories and the range of variation within and between your categories' (2014: 213)

She adds that if we define saturation as merely 'nothing new happening' in our data then we may arrive at a false conclusion. Saturation is not the same as the repetition of

events or stories in the data; the acid test is whether categories are rich and have conceptual depth. She also warns that if one reaches saturation too quickly it is likely that the analysis lacks criticality and complexity (2014: 215).

This argument that saturation may not be as obvious as it first seems, and that it must not be rushed, presents a challenge for the researcher. If it is more sophisticated and nuanced than expected and if is to achieve certain standards that take time to accumulate, then what are the processes that the researcher should engage in over time and how is one to gauge the richness or completeness of the conceptual understanding of the material?

While saturation may function as an important element within qualitative research, it is both problematic and a neglected area of the research process (Fusch and Ness, 2015: 1408).

Saturation – why it is problematic

There are four main difficulties identifiable in literature in relation to saturation: the misleading metaphorical picture; the problem of dislocation; process distraction; the need for quality and transparency.

The misleading metaphorical picture

To begin with, the term 'saturation' itself provides a misleading image; one which Dey (1999: 257) describes as an 'unfortunate metaphor'. Saturation suggests a point beyond which it is not possible to add anything further. To be fair, neither Charmaz (2014) nor Corbin and Strauss (2008) claim that saturation means 'completeness' in terms of understanding; they agree that it is very difficult to say that an entirely full conceptual understanding has been reached, a point beyond which it is impossible to go. Despite that, the metaphor has the potential to mislead and to encourage researchers to make claims beyond what is reasonable from the data. To illustrate this Thorne and Darbyshire (2005: 1108), with tongue in cheek, refer to saturation as 'the wet diaper', but the serious point they make is that the term is used to provide a false sense of completeness and an oversimplification of the data. In their particular field of health research they state: 'Despite health disciplines whose logic is invested in a theory of infinite possible variance of the inherent complexities involved, the saturation claim is often invoked as a convenient stopping point.' (2005: 1108)

A more appropriate way to define the point at which a researcher stops may be something like 'conceptual density' or 'conceptual depth'. To reach conceptual density is not to reach a final limit, beyond which it is impossible to achieve new insights, but it is to reach a *sufficient depth* of understanding that can allow the researcher to theorise. Dey (1999: 257) coins the phrase 'theoretical sufficiency' to describe this.

Dislocation

O'Reilly and Parker (2012) suggest that a key reason for the problems associated with saturation in qualitative research is the dislocation of the concept from its natural home in the grounded theory method. In grounded theory saturation is one element of theoretical sampling, which in itself has two core aspects. One is the iterative process of data

collection and anlaysis outlined above but another is the focus upon theoretical concepts in the data. Charmaz describes this as developing categories at an 'abstract and general level' (2014: 214). Similarly, Corbin and Strauss (2008: 75) say it is a process wherein the researcher looks at the 'property and dimensional level' of data. Saturation then is the point at which the researcher understands the theoretical categories to be sufficiently 'rich' and 'thick' (Fusch and Ness, 2015: 1409). Where the concept of saturation is applied in other contexts, there is the possibility that researchers mistake the repetition of events and incidents in the data as evidence of saturation (Charmaz, 2014: 213; Morse, 2015: 587). Kerr et al. (2010: 277) also believe that the problems associated with saturation are heightened when it is removed from the context of theoretical sampling, in which it was first developed, so different research designs may require different ways of dealing with saturation.

Process distraction

Because of the lack of clarity around the process of testing for saturation, there is potential for anxiety or uncertainty over what to do and how to do it. This preoccupation can have the effect of distracting the researcher from the fundamental tasks of building familiarity with the data and analyzing the complex and rich meanings within it. Piantanida et al. (2004: 332) use the phrase 'functional drift' to label this issue. In their own research they described it as: 'the pull toward specific, 'how-to' techniques that mimic the precision of post-positivist scientific methods'. In relation to saturation, it took them time to realise that 'the conscientious acquisition of texts' was less important than 'the quality of the researcher's interpretations of these texts.' (Piantanida et al., 2004: 337)

This over-emphasis upon procedure and a lack of familiarity with the data can be exacerbated through the use of Computer Assisted Qualitative Data Analysis Software (CAQDAS) according to Charmaz (2014). The ability to search and categorise quickly and in semi-automated ways can tempt the researcher to produce results too quickly without having conducted the analysis in a manner that is sufficiently methodical and detailed.

The need for quality and transparency

The need to adhere to sound methodological practice in qualitative methods if the work is to be considered valid, trustworthy and of sufficient quality has, over time, become well established (Flick, 2014; Miles and Huberman, 1994). This requires transparency and clarity in the process as well as provision of evidence in how judgments are made. Working within a qualitative research processes of Patient Reported Outcomes (PRO), Kerr et al. (2010: 271) explain that evidence of saturation is 'one of the most important aspects of qualitative methodological rigor' which their board[1] requires. Yet several scholars have identified a significant issue with how evidence of saturation is provided in research studies (Bowen, 2008; Caelli et al., 2003). In a survey of fifteen papers claiming to have achieved data saturation during their research, Francis et al. (2010: 1230) discovered that none had provided evidence of how saturation was achieved.

It is incumbent on a researcher, therefore, to not only take care in how decisions are made around reaching saturation but in how these are reported within the research. This is important because 'failure to reach data saturation has an impact on the quality of the research conducted and hampers content validity.' (Fusch and Ness, 2015: 1408).

Some responses to the problems with saturation

Given the importance of the issue of saturation in the qualitative research community, O'Reilly and Parker (2012: 196) have urged other researchers to engage with the problems, and there is some evidence that others have been bringing forward solutions. Responses to the problems of saturation show efforts to systematize, clarify and align.

In an attempt to systematize and clarify the process of reaching saturation, Brod et al. (2009: 1268ff) developed a 'saturation grid'. The method tabulates theoretical categories against data sources, which are listed in chronological order. The spaces on the grid are populated with the sub-categories (concepts) corresponding to the data source where they first emerged. When a data source produces no new sub-categories, then the core categories are said to be saturated. While there is a logic to this method it is likely to emphasise breadth over depth and doesn't provide a way to show relationships between categories (Kerr et al., 2010: 276).

In order to address these shortcomings, Kerr et al. (2010) developed the practice of 'annotated codebooks' alongside saturation grids to provide greater structure, transparency and, most importantly, evidence of depth in the analysis. The codebook is a table in which each code is explicated through: a brief definition; a full definition; an explanation of when to use and when not to use; and illustrated with a quotation from the data. They advocate that, used in combination, codebooks and saturation tables can provide substantial proof that saturation has been achieved.

In the case of both these studies it is interesting to note that as well as demonstrating a desire to systematize and clarify saturation they are also attempting to align it with a particular research design, patient reported outcomes (PRO). By contrast, Morse et al. (2014) have developed a saturation process particular to a PPGIS (public participation geographic information systems) qualitative study. In their article they identify a method whereby the general concept of saturation can be applied more meaningfully to their particular research design by plotting a graph of the number of samples by the total number of different responses. 'As the number of samples increases, fewer new responses are acquired, until the total number of different responses reaches an asymptote, at which point saturation of responses has been reached.' (Morse et al., 2014: 562)

What these suggest is that there is truth in the assertion that 'there is no one-size-fits-all method to reach data saturation' (Fucsh and Ness, 2015: 1409). Or as Kerr et al. (2010: 277) put it, saturation is not an 'atheoretical' generic research tool that can be applied in any qualitative research design. What is important is that there is alignment between research design and the saturation process employed.

To conclude this section, we can say that while saturation remains a commonly used tool within qualitative research, significant problems remain in relation to its definition and the processes associated with it. Attempts have been made to address aspects of these issues and these have focused upon the need to systematize, clarify and align saturation

within specific research designs. Keeping this discussion in mind the research undertaken for this present study sought to develop a response to the problems of saturation within a grounded theory approach. Given the fundamental importance of conceptual depth noted by grounded theorists above a method of saturation that employed the use of conceptual depth criteria was developed, and this is outlined below.

Development of conceptual depth criteria in a research context

All of the difficulties with reaching saturation were very much in mind during one grounded theory based piece of research which will now be discussed. The focus of the study was on investigating processes of sharing and collaboration between Post-Primary schools in Northern Ireland (Nelson, 2013). The schools in the region are largely separated along religious lines, with a small percentage of the pupil population (7%) attending purposely religiously-mixed schools. In a region which has had a history of conflict based around cultural and religious difference, sharing and collaboration between schools is not therefore straightforward. The research sought to explore the sharing activities of teachers when schools of different type chose to work together.

Agreeing with those who see the term 'saturation' as misleading, the term 'conceptual depth' was used instead. Flowing from this choice was the need, then, to establish a method for establishing *sufficiency* of conceptual depth. At the same time the criteria had to address the specific problems with saturation noted above. Building upon the responses developed by others in relation to developing an approach that was systematic, clear and aligned to the research design and keeping in mind broad criteria for quality in grounded theory research (Charmaz, 2014: 337ff; Corbin and Strauss, 2008: 305–309), a set of 'conceptual depth criteria' were established. A detailed explanation of each criteria is offered below but in brief they can be summarized as follows:

1. A wide **range** of evidence can be drawn from the data to illustrate the concepts.
2. The concepts must be demonstrably part of a rich network of concepts and themes in the data within which there are **complex** connections
3. **Subtlety** in the concepts is understood by the researcher and used constructively to articulate the richness in its meaning.
4. The concepts have **resonance** with existing literature in the area being investigated.
5. The concepts, as part of a wider analytic story, stand up to testing for external **validity.**

Criterion one – range

The range criterion requires evidence of multiple instances in the data which illustrate the conceptual categories. These instances are, effectively, the building blocks upon which a category is built and therefore the better the foundation the stronger the support for the category. In this regard range might be extended from meaning a range of quotations from different interviews to meaning a range of examples from different data sources.

The range criterion is likely to be the most easily satisfied of the criteria as the detailed nature of line by line coding produces multiple instances of codes (Urquhart, 2013: 159)

and these are easily quantifiable using computer aided qualitative data analysis software (CAQDAS). In my own study the use of MaxQDA software proved advantageous in producing evidence of code frequency, and code examples. Yet, despite the obvious assistance the software provided I was conscious that the use of technology in this situation had to be treated with caution for it tempts the researcher to equate sufficient range (in codes and data types) with high frequency. Rather, the focus had to be kept upon range in relation to meaning rather than in the frequency of occurrences (Morse, 2015: 587). For example, where a category was established, such as 'ethos' the range criterion raised a question as to whether there was a sufficient breadth of codes for the conceptual category to be meaningful. Or where a category was defined by competing perspectives such as 'children versus institutions' the issue of range raised a question of whether all positions on the issue were sufficiently represented. One tool which helped to establish the sufficiency of range was a positional map. The purpose of a positional map is to 'lay out major positions taken in the data on major discursive issues therein – topics of focus, concern and often but not always contestation.' (Clarke, 2005: 127). These have two particular advantages in that they force the researcher to look 'beyond binaries' and to 'hear silences'. In other words, they assist in understanding complexities and in identifying where, at times, there may shortcomings in the data range, that is, gaps, silences or previously unidentified positions.

The example, Figure 1, uses a number of quotations from the data to illustrate different positions expressed on a contentious question: when considering the extent to which separate schools should collaborate what takes priority – the needs of children or of schools? This issue was often articulated as a dualism when, in fact, the use of a positional map helped to identify shades of difference across the issue as well as drawing attention to a gap – no-one articulated a win-win scenario that would be good for children and schools. In other words, used as part of criteria for conceptual depth it was possible to see that the range test had not been fully satisfied. Clearly, it was important to ascertain whether further shades of meaning on this issue were present in the field of study.

When coming to a decision about conceptual depth this criterion is fulfilled by ensuring that multiple examples of concepts in the data can be provided and evidenced across a range of sources, but achieving a reasonable frequency of codes is not in itself an indicator of range. Instead range must be judged in relation to the meanings contained within the conceptual categories and, inevitably, that means that it is not understood in isolation from the following two criteria in particular: complexity and subtlety.

Criterion two – complexity

The complexity criterion demands that concepts must be demonstrably part of a rich network of other concepts and themes in the data within which there are complex connections. One of the advantages of grounded theory is that a wide variety of ways have been developed to assist in this, from coding-trees to maps, matrices and diagrams. Again, many of these can be generated through the use of CAQDAS although from my own experience I have found that a combination of methods of production is valuable. At an early stage hand-sketched diagrams or matrices helped me to 'play' with the concepts assisting me in sorting codes, investigating comparisons, identifying categories and, most importantly, understanding the connections between concepts. Only at a later stage,

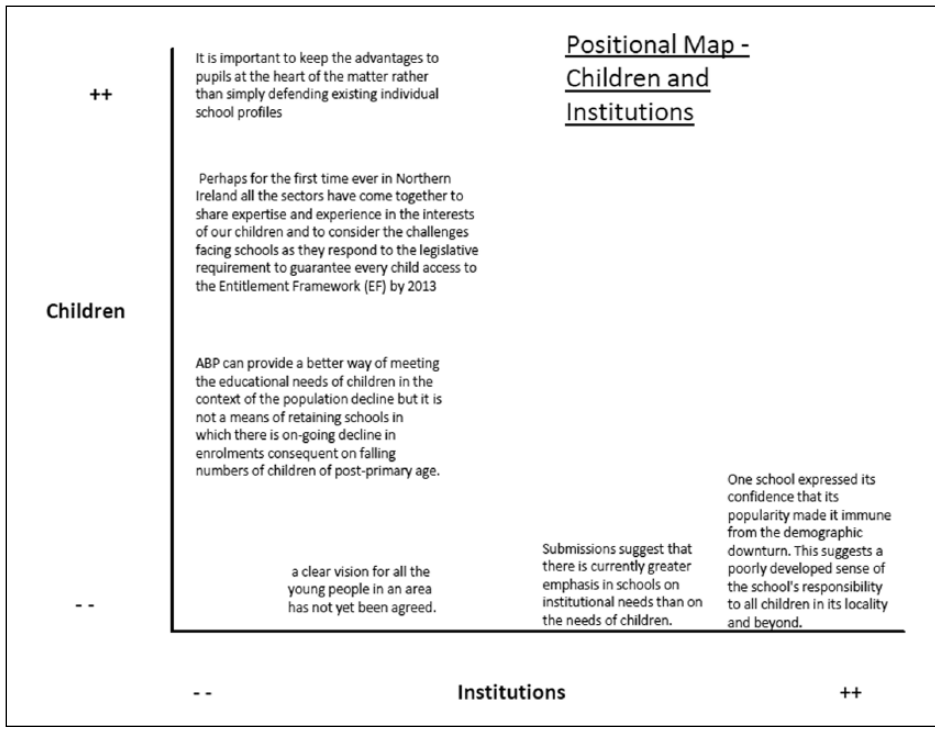

Figure 1. Positional map.

when a significant amount of coding had been carried out and conceptual categories were becoming defined, were more sophisticated means used to produce the drawings.

The use of diagrams has been a regular feature in many grounded theory books and articles (Corbin and Strauss, 2008; Mills et al., 2007), although Adele Clarke has probably exploited their use most fully in her version of grounded theory, situational analysis (Clarke, 2005). She believes they maximize the reflexive thrust of grounded theory helping us 'to see things afresh', to make 'assemblages and connections' and are devices for 'analyzing relationality'.

In my research on sharing processes between schools, use was made of positional maps, social worlds diagrams, situational maps and matrices (Clarke, 2005) as well as CAQDAS generated concept maps. One example, Figure 2, shows an analysis of how teachers integrate collaborative activity into their work. Teachers were faced with the dilemma of how to integrate the collaborative work into their existing workload and they would juggle this by carrying out some of the work through extra-curricular activities (clubs, societies and special off-timetable days) and some within existing curriculum work. Examples of this in the data were coded as 'extra vs core' but it was also interesting to note that decisions about how to manage these activities were also based upon cost/benefit judgments in terms of time and effort. Representing this in diagrammatic form added to the analysis by illustrating how teachers attempt to turn costs into benefits by making a 'fitting' argument (the positions taken in quadrant A).

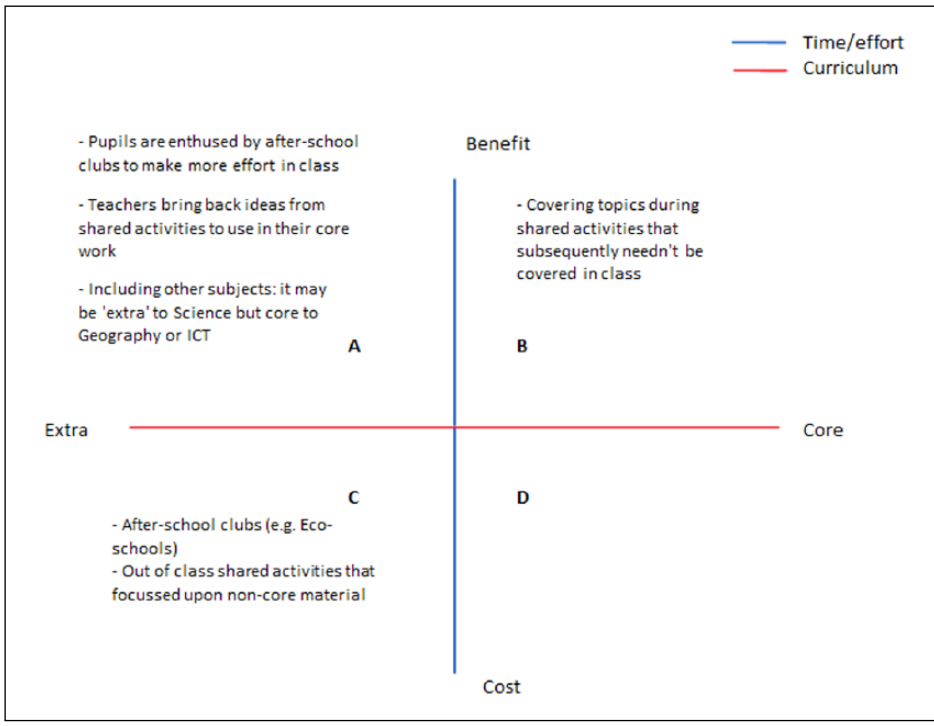

Figure 2. Curriculum fit matrix.

The benefit of maps and diagrams is that they can achieve dual outcomes of helping the researcher to represent complexity and address the issues of transparency and credibility (Buckley and Waring, 2013; Wu and Beaunae, 2012). Only a small number of maps and diagrams generated during the analysis may make it into a final publication, but utilizing them throughout the analysis can prove crucial in improving the sophistication of the analysis and be of critical importance when working towards a theoretical understanding of concepts.

Criterion three – subtlety

Subtlety is developed as a result of the comparative method whereby a researcher compares the instances of the same codes and asks how they are similar or different. This teasing out of meaning is central to grounded theory method and can help the researcher to determine shades of difference in the meaning of different words or concepts and may also help identify ambiguities.

In my own study, for example, a common theme was 'our children'. Teachers often used the phrase when justifying certain decisions or when highlighting the benefits of particular curricular interventions ('we do it for our children'). Interestingly, through comparison of the occurrences of 'our children' it could be seen that the range of meanings included: class or year groups; children belonging to one school; children from different schools who were participating in inter-school activities; all children from a geographical region. In most

cases the phrase may not have been specified by the teacher as the context of its use was assumed to define it, but when the code 'our children' was understood as part of a higher level category, 'identity', the multiple meanings of the phrase did highlight ambiguity in the teachers' representation of the identities of young people.

Understanding subtlety and ambiguity of concepts was facilitated in particular through memo writing. Memo-writing is encouraged as a core activity for the grounded theory researcher (Charmaz, 2014: ch.7). In my research, to begin with, memos were elaborations of my codes, usually just a few sentences, but as the analysis progressed they were revisited, expanded, shaped, merged with others and generally used to develop descriptions of full-bodied concepts. In taking the time to articulate concepts in this way I was challenged to come to terms with the contradictions, nuances and ambiguities within them. An extract from a memo on an 'in-vivo' code (where a word or phrase used by participants is adopted as a code by the researcher) is included below (Figure 3) to illustrate how the memo-writing process provided space for reflection and an environment for the exploration of nuance, subtlety and ambiguity. The code label was 'No issues' - a phrase used by respondents in interviews when they were asked to comment on any challenges that had arisen around community relations' issues during shared educational activities.

For the subtlety criteria to be met the researcher must check that they have taken time to interrogate the language used by participants in the research and the language used in the conceptual categories generated during the data analysis. As with the range criterion above there is no particular number of occasions where this criterion must be satisfied, but if memo writing is a feature of the analysis process at every stage then a key task of the memo-writing process should be to expand the meaning of conceptual categories in response to questions which explore the hermeneutics of the language used. Where conceptual language is regarded as unproblematic and one dimensional then subtlety is unlikely to be present, however, where conceptual language is understood as rich, ambiguous and multi-dimensional then subtlety will be evident.

Criterion four – resonance

For many grounded theory practitioners, choices about theoretical sampling should be solely dependent upon what emerges from the data. But what if what emerges from the data is conceptually remote from any other ideas or theories in the academic literature? Surely, it has to raise serious questions in the mind of the researcher. If, however, there is resonance with existing literature (overlaps; similarities of language and metaphor etc.), albeit with variations and novelties, this can be used as further confirmation that a sufficient conceptual depth has been reached. Indeed, Urquhart believes it 'necessary and desirable' to relate an emergent theory to literature (Urquhart, 2013: 136).

The position of a literature review in grounded theory studies has been a source of controversy for some time (Dunne, 2011). Initially, Glaser and Strauss (1967) recommended abstention from literature prior to data collection and analysis in the hope that the researcher's thinking would not be contaminated by existing theories. Despite Glaser's continued insistence on this point (Glaser, 2002, 2012) there are few who achieve it or who support it (Charmaz, 2014). Strauss later altered his position arguing that to conduct a literature review prior to data collection is reasonable on the basis that

'No issues' memo

'There are no issues'. This could mean several things:

- No issues of difference come to the surface because there has been no risk-taking in terms of reconciliation work

- The schools can work together with no issues because of the extensive common ground that exists around curriculum collaboration and this focus upon academic knowledge is uncontested.

- Blindness to issues. Perhaps there are issues but the teachers exercise selective blindness and choose not to see them or report them

- Avoidance of diversity/reconciliation issues. This is a similar to the first point but on a different level. Rather than limited engagement of community relations' issues there is active avoidance of raising any issues which may be perceived as difficult

- I am asking the wrong questions. Perhaps there are issues but the questions that I am asking are not facilitating a conversation about them

- I am looking for something that doesn't exist or, at least, is not recognisable to those participating in the collaboration. In other words, there are no issues!

Figure 3. No issues memo.

theoretical knowledge is not the same as theoretical bias, although he offers no substantiation of this assertion and, arguably, it is a view that ignores what has been generally accepted in the field of psychology for some time: our judgments are likely to be significantly influenced by the ease with which we can bring relevant evidence to mind. This is known as the 'availability heuristic' (Tversky and Kahneman, 1974) and, it raises a genuine challenge to Strauss' argument by implying that knowledge or theoretical perspectives recently acquired through a literature review may be likely to influence (or bias) a researcher simply due to the fact that recent exposure to them means that they will easily come to mind. So where does this leave the grounded theory researcher? Should they attempt to ignore literature, a proposal which is likely to be artificial and unrealistic, or embrace literature with a näive assumption that they can know about a range of theoretical perspectives but avoid them having any influence on their study? Over time a pragmatic perspective has emerged in a number of studies (Bowen, 2006; McGhee et al., 2007): firstly is an acknowledgement that no researcher is a *tabula rasa*, with or without a literature review, so, secondly, what is required is an informed self-awareness of the positions that one brings to the data analysis and the ability to 'turn back' to appraise their effects (McGhee et al., 2007).

In my research I engaged with literature at the outset, even before I had fully settled upon a research methodology, but once I decided to pursue a grounded theory methodology I chose to mitigate the effects of this in two ways: I used 'reflective memos' to help me identify my own bias and I avoided literature during the first two phases of theoretical sampling.

These actions gave me a certain freedom and freshness in my analytical approach, but at times I also experienced some anxiety in not knowing if my analysis was proceeding in an appropriate and credible direction. Ultimately, I felt that the theoretical propositions which I was developing needed to be tested against other perspectives and that this

would be a critical test of their depth. This proved to be a reasonable and appropriate choice and not remote from what others have recommended (Bowen, 2006, 2008).

Having conducted several rounds of data collection and analysis I used the criterion of literature to test the sufficient conceptual depth of my emerging theory. My analysis had given me new theoretical routes to explore and meeting this new literature at this stage was vitally important for it confirmed that the analytical route I had embarked upon was a reasonable one but it also suggested that more theoretical sampling was needed if the results were to be scaled up to an appropriate level. As a result, it provided me with the confidence to proceed with my analysis in a more informed and critical manner.

Criterion five – validity

If the results from research are to be of any value they must be credible and generate learning which will be useful to others (Rossman and Rallis, 2012). It is reasonable, therefore, for those reading research results to ask questions about the validity or reliability of the findings. Flick (2014: chapter 29) has noted that these are particularly challenging questions for authors of qualitative studies and Dey asserts that validation is particularly difficult for grounded theory researchers, and a significant casualty from the practice of theoretical sampling (Dey, 2007: 84). Yet, despite this, funding authorities and journal editorial boards are increasingly seeking evidence of validity in relation to saturation in qualitative studies (Kerr et al., 2010; O'Reilly and Parker, 2012).

In opening up these questions of validity and reliability it is important to acknowledge the wide debate which exists around their definition and application (Flick, 2014). Not only can the terms be used in ways within the realms of qualitative research which are overlapping but they are also inter-mingled with other evaluative terms such as generalizability, verifiability and dependability which each bring their own etymological and semantic baggage. Maxwell's typology of validity (1992) provides much needed clarity by defining five types of validity: descriptive; interpretive; theoretical; generalizable and evaluative. And a primary difference among these is that the first three refer to validity in relation to the methods, procedures and presentation of data collection and analysis (how can we be sure the account provided by the researcher is to be believed?) while the latter two refer to issues beyond the research process itself – other social contexts and the moral framework brought to the study by the researcher (to what extent do these results provide outcomes which stand up to scrutiny in the world beyond the immediate context of the research? And to what extent are any value judgments made by the researcher a legitimate part of the analysis?). This distinction between the first type (what we could call 'internal validity') from the second type (what we could call 'external validity') is important for it shows that Dey's assertion that validity is a casualty of theoretical sampling requires qualification. Theoretical sampling does have abundant evidence to offer in terms of 'internal validation', as has been shown in discussion of criteria 1–3 above; the use of maps, memos and the constant comparative method are explicit means for demonstrating credible and reliable processes in the research. They can help to satisfy what Bowen (2008: 148) describes as the 'trustworthiness requirement' of qualitative research. However, the issue of 'external validity' is more challenging, for it requires answers to more difficult questions: can a grounded theory be applied beyond its immediate context? Does it have generalizable or universal elements?

Grounded theory researchers aspire to 'general' theoretical concepts (Glaser and Strauss, 1967) but this is not understood as meaning that outcomes from grounded theory studies possess 'universality'. So what does it mean to say that the outcomes must make sense in a general way? For Corbin and Strauss (2008) this means that the findings should possess applicability to those in the field of study or ones which are similar. In other words, grounded theory researchers must avoid the temptation to stay inward facing in their research. This inwardness could be characterised as being content for the findings to merely state detailed descriptions of the field of study, or to be articulated in such specialised language that the results only make sense to the researcher or a very small group of academics. Instead, a study which aims for applicability should be outward facing in two ways: On one hand, the aim should be for the findings to be understood in conceptual terms which raises the level of analysis above technical description to more general themes. On the other hand, the outcomes should be expressed in terms which can be understood by those who have familiarity with the social context, or ones broadly similar, about which the results speak.

Aware of this distinction between internal and external validity I sought to satisfy this fifth criterion for validity in my own research by a focus on 'external validity' and applicability. The importance of an outside view has already been mentioned in relation to testing a development theory against literature (criteria 4) and a similar attempt to make sense of the findings from a position outside of the immediate field of research was employed here. Would the concepts already developed have applicability in a setting similar to the original?

I chose to check the emerging theoretical findings against data from teachers in other schools not previously used as sites for data collection. The original schools in my field of research were part of one sharing cluster, but there were a dozen other clusters doing broadly similar work so it was possible to use these as test sites for applicability. The use of new sites is suggested by Glaser and Strauss (1967) and Corbin and Strauss (2008: 155) as a means of widening the scope of the theory, something Urquhart (2013) also supports and wonders why the strategy isn't taken up by more researchers using grounded theory methods.

In practical terms this meant coding of further interviews and observations, but not in the same detail as the early analysis. At this stage a theoretical sensitivity had been developed and it was possible to use the emerging themes as part of the coding process. In terms of applicability it was important to ascertain which of the main themes were useful and which were not and to keep an open mind to whether new themes were emerging. In this way, some comparison was possible across the sites in order to determine whether there was general applicability of the themes and the emerging theoretical perspectives.

Applying the criteria

In my own research example described above I employed the conceptual depth criteria twice. After the first occasion I had taken 27 'slices of data' (9 interviews; 7 observations; 11 documentary sources) but felt unable to satisfy the sufficiency criteria for all the categories. The criteria had helped me to identify that the 'resonance' and 'validity' categories were weak and, while there was better evidence that the other criteria were being developed, there was still room for improvement. Yet, I was also aware that I had

arrived at this judgment instinctively and in a way that could be difficult to convey to others. Reviewing the theoretical sampling process one more time I felt a more precise measurement of 'sufficiency' was possible through the use of general descriptors ('few', 'abundant, 'weak', 'strong' etc.) and a three-point scale (see Table 1). What had been articulated largely as intuitive knowledge about the data was now more clearly defined and transparent.

After further rounds of theoretical sampling which included a return to the field for further interviews and observations as well as returning to core documentary sources and reviewing what I was now able to identify as relevant literature, a second test using the conceptual depth criteria was conducted. The scores from the two tests for conceptual depth could be compared (see Table 2) and, at this stage, it was felt that the sufficiency criteria had been met and, while more work was still needed to refine and scale-up the theory, theoretical sampling could cease. Of course, a scale of this kind requires testing in a variety of research contexts to prove that it has value but it may assist qualitative researchers in providing a stronger evidence-base for their decisions in relation to theoretical sampling and in establishing conceptual depth.

Table 1. Conceptual depth scale.

Criteria (with sources of evidence)	Low (1)	Medium (2)	High (3)
Range (e.g. frequency and variety of codes; multiplicity of data sources	Few examples to support concepts. Only a single data-type	→	Abundant examples to support concepts. Multiple data-types
Complexity (e.g. coding trees; positional maps; matrices)	Descriptive codes; simple or basic connections between codes; low level analysis	→	Sophisticated networks; abstract conceptual categories which synthesise a range of codes and concepts
Subtlety (e.g. memos; social worlds diagrams)	Conceptual language is regarded as unproblematic and one dimensional	→	Conceptual language is understood as rich, ambiguous and multi-dimensional
Resonance (literature)	Weak resonance; emerging theory is remote from existing literature and theoretical frameworks	→	Strong resonance; emerging theory makes sense alongside existing literature; there are correlations with other theoretical frameworks, albeit with variations and novelties
Validity (e.g. applicability test)	Low level theorising and inward facing; the findings have limited application to the research participants or those familiar with similar contexts.	→	Abstract level theorising and outward facing; the findings make sense to those in the social context of the research, or ones broadly similar.

Table 2. Test scores.

	Test 1	Test 2
Range	2	3
Complexity	2	3
Subtlety	2	3
Resonance	1	2
Validity	1	2
Total score	**8**	**13**

Conclusions

This article has sought to investigate the concept of 'saturation' which is an important aspect of qualitative research. It has identified problems arising from the term itself as well as the application of it during the research process. Having considered the issues broadly, it was agreed that there may not be a one-size-fits-all response to the problems associated with saturation, but solutions need to be developed in alignment with specific research designs. In the case of this research, solutions were brought forward in alignment with a grounded theory approach.

The phrase 'conceptual depth' was suggested as an alternative to the term 'saturation' and the process of determining conceptual depth was addressed using five criteria. It is concluded that, limitations notwithstanding, the use of conceptual depth criteria can aid the researcher in judging whether conceptual categories are sufficiently robust and capable of resourcing the theorising stage of the grounded theory method.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Note

1. US Department of Health and Human Services Food and Drug Administration (FDA)

References

Bowen GA (2006) Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods* 5(3): 1–9.

Bowen GA (2008) Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research* 8(1): 137–152.

Brod M, Tesler LE and Christensen TL (2009) Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research* 18 (9): 1263–1278.

Buckley CA and Waring MJ (2013) Using diagrams to support the research process: examples from grounded theory. *Qualitative Research* 13(2): 148–172.

Caelli K, Ray L and Mill J (2003) 'Clear as mud': toward greater clarity in generic qualitative research. *International Journal of Qualitative Methods* 2(2): 1–24.

Charmaz K (2014) *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis* (second edition). London: SAGE.

Clarke A (2005) *Situational Analysis: Grounded Theory after the Postmodern Turn*. London: SAGE.

Corbin J and Strauss AK (2008) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: SAGE.

Dey I (1999) *Grounding Grounded Theory*. San Francisco, CA: Academic Press.

Dey I (2007) Grounded Theory In: Seale C, Gobo G, Gubrium JF and Silverman D (eds) *Qualitative Research Practice*. London: SAGE, pp. 80–94.

Dunne C (2011) The place of the literature review in grounded theory research. *International Journal of Social Research Methodology* 14(2): 111–24.

Flick U (2014) *An Introduction to Qualitative Research* (fifth edition). London: SAGE.

Francis JJ, Johnston M, Robertson C, Glidewell L, Entwistle V, Eccles MP and Grimshaw JM (2010) What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health* 25(10): 1229–1245.

Fusch PI and Ness LR (2015) Are we there yet? Data saturation in qualitative research. *Qualitative Report* 20(9): 1408–1416.

Grimshaw JM (2010) What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health* 25(10): 1229–1245.

Glaser BD and Strauss AK (1967) *The Discovery of Grounded Theory*. Chicago, IL: Aldine.

Glaser BD (2002) Conceptualization: on theory and theorizing using grounded theory. *International Journal of Qualitative Methods* 1(2): 23–38.

Glaser BD (2012) No preconceptions: the dictum. *Grounded Theory Review* 11(2): 1–6.

Guest G, Bunce A and Johnson L (2006) How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18(1): 59–82.

Kerr C, Nixon A and Wild D (2010) Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. *Expert Review of Pharmacoeconomics and Outcomes Research* 10: 269–281.

Mason M (2010) Sample size and saturation in PhD studies using qualitative interviews. *Forum: Qualitative Social Research* 11(3): Art. 8.

Maxwell J (1992). Understanding and validity in qualitative research. *Harvard Educational Review* 62(3): 279–300.

McGhee G, Marland GR and Atkinson J (2007) Grounded theory research: literature reviewing and reflexivity. *Journal of Advanced Nursing* 60(3): 334–342.

Miles MB and Huberman AM (1994) *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: SAGE.

Mills J, Chapman Y, Bonner A and Francis K (2007) Grounded theory: a methodological spiral from positivism to postmodernism. *Journal of Advanced Nursing* 58(1): 72–79.

Morse JM (2015) 'Data were saturated . . .'. *Qualitative Health Research* 25(5): 587–588.

Morse JM (1995) The significance of saturation. *Qualitative Health Research* 5(2): 147–149.

Morse WC, Lowery DR and Steury T (2014) Exploring saturation of themes and spatial locations in qualitative public participation geographic information systems research. *Society and Natural Resources* 27(5): 557–571.

Nelson J (2013) Common education and separate schools: a study of sharing education in Northern Ireland using a grounded theory methodology. PhD Thesis: Queen's University Belfast.

O'Reilly M and Parker N (2012) 'Unsatisfactory saturation': a critical exploration of the notion of saturated sample sizes in qualitative research. *Qualitative Research* 13(2): 190–197.

Piantanida M, Tananis CA and Grubs RE (2004) Generating grounded theory of/for educational practice: the journey of three epistemorphs. *International Journal of Qualitative Studies in Education* 17(3): 325–346.

Rossman GB and Rallis SF (2012) *Learning in the Field: An Introduction to Qualitative Research*. London: SAGE.

Thorne S and Darbyshire P (2005) Land mines in the field : the craft of qualitative health research. *Qualitative Health Research* 15(8): 1105–1113.

Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–1131.

Urquhart C (2013) *Grounded Theory for Qualitative Research: A Practical Guide*. London: SAGE.

Wu C-HV and Beaunae C (2012) Personal reflections on cautions and considerations for navigating the path of grounded theory doctoral theses and dissertations: a long walk through a dark forest. *International Journal of Social Research Methodology* 17(3): 249–265.

Author biography

James Nelson is Lecturer in Education, coordinator of the PGCE in Religious Education and programme coordinator of the Masters in Education Studies at Queen's University Belfast.