Staying True to the Data: A Visual and Quantitative Approach to Showcase Coding Rigour and Theoretical Saturation

Ilse Doyer
Department of Industrial and Systems Engineering, University of Pretoria
Michael K. Ayomoh
Department of Industrial and Systems Engineering, University of Pretoria

Abstract

This paper describes a data-based and transparent approach that was followed during classic grounded theory data analysis to ensure rigor and theoretical saturation. Firstly, a visual approach was followed to ensure that the theory discovery stayed as close to the story told by the data set as possible, thus ensuring both rigor and accuracy. Secondly, process data in the form of a coding rate time study and theoretical contribution rate, were tracked to monitor theoretical saturation objectively. The presented approach offers a novel approach to showcasing classic grounded theory rigor during the data coding process, as well as a way to prove that theoretical saturation was reached. The approach presented in this paper thus provides grounded theory researchers with a way in which to both ensure and defend the rigor and accuracy of their research.

Keywords: Classic grounded theory, theoretical saturation, sample size, time study, data-based approach, data analysis

Grounded theory (GT), first pioneered by Glaser and Strauss in the 1960s is a systematic, qualitative method of inductive, iterative, comparative and interactive data collection and analysis with the aim of theory building (Charmaz & Belgrave, 2012). It's data based theory-building approach was developed in response to the 'armchair' theory-building that had become prevalent at that time (Urquhart et al., 2010). It is often used in a field of study where little is known about the field and approaches data collection and analysis with an open mind so that theory formation is 'grounded' in (as opposed to uninformed about) the realities of the target domain (Goulding, 2002). A major distinguishing trait of GT is the fact that no hypothesis is formulated upfront, and limited to no research is done on the topic so that the researcher has as few preconceptions as possible when embarking on the research (Vander Linden & Palmieri, 2021). To achieve this end, classic GT differs from the other GT schools in that it completely avoids literature reviews until the very end of the analysis process (Kenny & Fourie, 2015) .

Qualitative research, in general, is often criticized for its lack of scientific rigor (Welch & Piekkari, 2017). A common criticism levelled against GT is the extent to which the research can really claim to be objectively representing the data without too much interpretation or extrapolation by the researchers tasked with coding and categorizing this data (Kenny & Fourie, 2015, p. 1274). Especially in classic GT, where the researcher is expected to maintain an objective stance to the coding of the data and stay as close to the story told by the data itself as possible in order to discover, as opposed to creating (as in Straussian GT) or constructing (as in constructivist GT), the grounded theory (Kenny & Fourie, 2015), ensuring rigor and transparency during the coding process is imperative to the credibility of the classic GT research process. This paper presents a transparent and data-based approach that creates both accountability and transparency during the data analysis phases and aims to enable classic GT researchers in proving and defending the rigor and accuracy of their research.

Interviews are the most common method of data collection in GT, providing a rich and nuanced data set from which to extract theory that is grounded in the data. In an analysis of 100 GT studies, Thomson (2011) found that the point of theoretical saturation can be affected by the scope of the research question, the sensitivity of the phenomena, and the ability of the researcher, but that a sample size of 30 interview participants is a good rule of thumb for GT theory building. Vander Linden and Palmieri (2021) state that the quality of the collected data, and the scope of the theory determine the sample size that will deliver theoretical saturation, and that principles should be adhered to rather than a rule of thumb but mention 9 – 30 participants to be generally accepted rule of thumb in GT research (Vander Linden & Palmieri, 2023). Thus, although sample size rules of thumb exist, scholars agree that the ultimate determinant is whether theoretical saturation has been reached, which depends on many different factors. Considering that the discernment of whether theoretical saturation has been reached lies in the hands of the researchers and reviewers, this important step may be susceptible to subjectivity, confirmation bias, and a lack of integrity or rigor.

This paper proposes and demonstrates how data can be used to monitor the theoretical saturation arch rather than depending on a rule of thumb for sample size that might not relate to the research scope or leaving the decision to the discernment of invested parties. This paper also demonstrates how visual transparency during theory building can keep researchers focused on staying true to the story told by the data and not take short cuts to theory building based on bias or superficial coding. The aim of the research is to make both an empirical and methodological contribution to the field of GT by presenting a GT case study that shows systematic and transparent model development, as well as how data were used as a more objective, real-time approach ensuring saturation.

Background to the Research Design and Execution

The topic of the research centered around the selection of improvement projects in operations improvement, thus essentially a decision sciences topic, but in a specific setting. Little research has been done on the praxis of project selection, even though whole fields of study are dedicated to the development of mathematical models and computerized decision support systems. The research question explored in this case study was thus simply the open-ended: "How are operations improvement projects selected in practice?"

Ten unstructured interviews were held with seasoned operations improvement practitioners, each having more than 25 years of first-hand operations improvement experience. Together, the participants could report on 328 years of observing the selection of improvement projects in operations and provided a rich data set collected during the interviews that lasted between 60 and 120 minutes each.

Ethical clearance was sought from the affiliated research institution's ethical committee before data collection commenced and each informant submitted a signed informed consent letter before the interviews, which stated that participation was voluntary and that responses would be treated confidentially. Participants were numbered, and a single letter was used to help the researchers recall the context of the first-order code more easily. No company-specific information was noted.

The interviews were conducted online and recorded. The recordings were used to capture 322 lines of discrete statements that could be analyzed. Microsoft Excel was used throughout the analysis for its transparency, meaning that the researcher is constantly able to scroll, move, copy, paste, and reference all fields of the data.

The data analysis was conducted twice, one year apart, which, in retrospect, enabled what Urquhart et al. (2010) described as “living with the data for a long time”. The initial analysis, unfortunately not tracked as demonstrated in this paper for the second round of data analysis, revealed that the topic being studied, project selection, is not a discrete process step, but that it, in practice, often overlaps with other project phases such as project identification. During the interviews, interview participants, in spite of their significant experience and cognitive capabilities, kept deviating from pure project selection to describe other processes. This created confusion in the first-order analysis phase until a first “a-ha moment” allowed the researchers to recognize this a pattern within the data, which constituted a deviation between praxis and how the topic is taught in business schools and treated in literature. .

The second “a-ha moment” was that participants described how projects are selected just as much as they described what can go wrong during project selection. This meant that, staying true to the story told by the target domain, that project selection is as much about the methods employed as it is about the phenomena that influence the effectiveness of these methods. This led to the identification of the data categories of method, as well as decision-making impediments and enhancements (DMI/DMEs).

This initial round of data analysis followed many of the rigor principles promoted in GT literature but was not documented as rigorously as the second round, making the findings hard to publish in journals cynical about the weaknesses presented by superficial GT coding. Hence, it was decided to go through the data set once again to meticulously document the train of thought leading to the findings. This was both grueling and satisfying work as additional findings were made and a deeper analysis was done that during the first round. The researchers speculate that the increased depth of insight had as much to do with the fact that the researchers had had time to process their initial exposure to the shape of the data, as with the systematic rigor applied for the documentation of the process, but that the rigor of meticulously working through each line of data to document what was done with it, had a significant and rewarding effect on the theory building process.

To truly do justice to the classic GT process, the second round of data analysis was approached with a determination to stay as true to the wording and meaning of the participants' statements as possible. This determination was rewarded with many new insights, as well as corrections to previous perceptions. The following seven categories were identified, iteratively, during the course of the first-order analysis of the first third of the 322 lines: Methods, criteria, circumstances, decision-making impediments (DMIs), decision-making enhancements (DMEs), process flow relationships, and researcher notes and questions, which were kept and tracked separately so as not to influence the classic GT process.

Once the first-order coding had stabilized around the listed categories, a basic process flow started to be constructed visually on a wall in the main author's office, eventually containing another 11 categories or process steps. This model developed significantly during iterative coding of the first three participants' statements, then slowed down during the coding of the next two participants' statements, and completely plateaued throughout the next five participants' coding, indicating that theoretical saturation had been reached as new data did not contribute any new theoretical insights (Hallberg, 2006; Urquhart, 2019).

Although the study by Thomson (2011) found that the average sample size in GT is 25, and recommended that 30 be regarded as a minimum for rigor, their study also found that theoretical sampling was influenced by the scope of the research question, the sensitivity of the phenom3na, and the ability of the researcher. In the case of the research described in this paper, section 3 will show that theoretical saturation had, in fact, been reached after five participants. Although it is difficult to compare the scope of the research question, or the sensitivity of the phenomenon being studied to that of other studies, it can be noted that the main author has also spent 20+ years in the operations improvement sphere in practice, making it easier to understand the data and recognize the structure of the phenomenon. Additionally, the year that elapsed between the two rounds of data analysis also added to the rate at which the dots could be connected during the second round of analysis.

Although the pace of model development plateaued after five participants' data was coded, the last five participants contributed further examples in each of the categories. Theoretical sampling then guided the researchers to rather test the completeness of the categories with a structured online survey to reach a wider pool of operations improvement practitioners that could expand the lists already captured within each of the data categories. The survey was completed by 182 participants, and only five additions were made, indicating that the ten interviews had indeed collected a rich data set. Survey participants were also asked to indicate whether they had observed the examples provided by the interviewees, thereby testing the validity of the examples included in the model.

The meticulous documentation of the second round of data analysis stimulated a meta-analysis of the coding and theory building process itself. The results of this coding process analysis are discussed next to demonstrate the arch of the GT data analysis.

Visual and Transparent Model Development

Although not the most important contribution of this paper, this discussion will start with an introduction to the visual model to provide a point of reference for the rest of the section. This discussion will also show how the model visually shows an adherence to the principles of rigor and accuracy of classic GT to ensure that the model development stays true to the inputs of the target domain, while also allowing for researcher insight.

Figure 1 shows the theoretical model with the white elements showing the main phases of the process being studied, the red arrows indicating process flow, and the color-coded squares representing examples of each of the data categories, explained in the legend to the diagram. Most importantly, as can be seen in the legend, special provision was made to separate researcher contributions from the rest of the model, which is solely built from participant input unless indicated with major or minor pink squares, or small red squares where a term was borrowed from known theory to summarize a participant statement. By visually highlighting researcher interaction, the visual model lends transparency to the GT coding process, keeping the researchers accountable to staying as close to the respondents' terms as possible.

The theoretical model is not the focus of this paper and will thus not be discussed in detail. Rather, the focus is on how rigor and theoretical saturation were tracked and visually audited to assure quality in the theory-building.

Figure 1

Final Theoretical Visualised "Wall" Model with Pink Blocks Indicating Researcher Insights or Questions

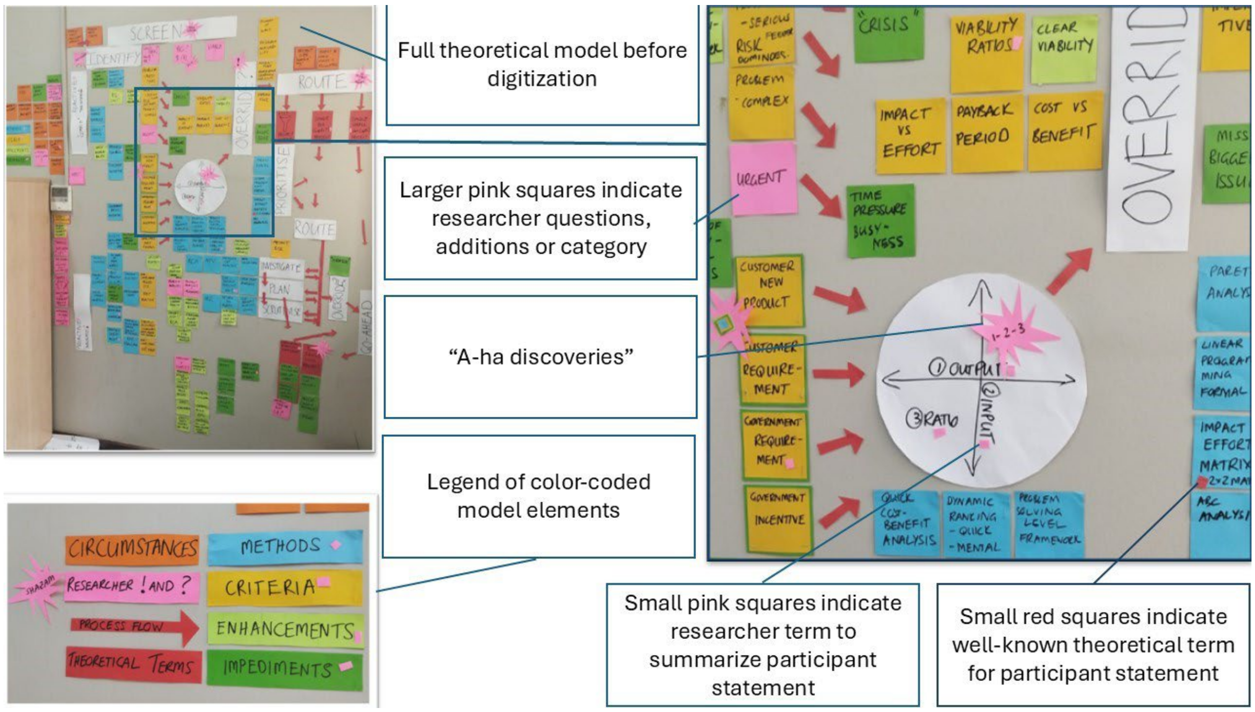


Figure 1 visualizes, and Table I analyses the level of objectivity of the analysis by indicating (Figure 1 pink elements) or counting (Table I) where researcher/literature contributions were made and where the model is built from interview or survey participation alone. The analysis in Table I shows that 82.9% of the model wording and concepts can be traced back to one or more participants' statements, meaning that the researchers contributed only 17.1% of the elements, with 16.9% being a theoretical/ structural contribution and only 0.7%, one single distinct element being added by the researchers, having been extrapolated from a list of criteria provided by the interview participants, which seemed to have one obvious omission. Although both the speed at which theoretical saturation was reached, and the seemingly low proportion of researcher contribution, the theoretical contribution of articulating the "anatomy" and "physiology" (Gioia et al., 2013, p. 24) of project selection in practice is a major leap forward from the oversimplified interpretation currently given to project selection by the existing body of knowledge.

In classic GT, researcher notes are encouraged to keep track of emergent theory questions as the coding of the data progresses (Vander Linden & Palmieri, 2023, p. 9), indicated by the eight (4.8% of all visual elements) larger pink squares on the visual model. In 18 elements, the participants' statements would not contain a specific noun that could be used as an element and would thus be summarized or interpreted by the researchers. These elements were, however, the conceptual contributions by the participants.

The clear indication of where the researchers deviated from the pure data set contributed by the interview and survey participants, allows the researchers to remain critical or open-minded of their own contributions, flagging them for consistent theoretical investigation.

Table 1

Analysis of Participant and Researcher Contributions to Model

Counted Visual Element	#	%
Total visual elements incl. notes and small indicators	166	
MINUS researcher questions/notes (some larger pink squares)	8	4.8%
MINUS participant concept, researcher term (small pink squares)	13	7.8%
MINUS participant concept, literature term (small red squares)	5	3.0%
Total model elements (sans researcher notes and labels)	140	
MINUS researcher theory contribution (white headings and some larger pink squares)	23	16.4%
MINUS researcher extrapolated example (one larger pink square)	1	0.7%
Pure participant contribution	116	82.9%

The transparency of the analysis provided by the visual model (Figure 1) is echoed in the Microsoft Excel spreadsheet where the analysis process was meticulously documented. The main analysis sheet contains the 322 lines of participant statements, which are then broken down into columns – one for each model category. Separate sheets were also created, one per model category, where, for example project selection method examples are listed and described by who, what, and how many questions, i.e. which participant(s) mentioned this method, how many participants mentioned this method and what did they say about it. The Microsoft Excel spreadsheet thus makes it possible to trace the model elements and structure to the original participant statements.

Monitoring Progress Towards Theoretical Saturation

As the researchers are Industrial Engineers, it naturally occurred to collect time study data as the analysis unfolded as there was a significant trend that reflected arch of theory development. The first order analysis saw several iterations to find order and patterns in the data. In fact, the first nine lines of data took 3.3 hours of iterative coding to find categories that could describe the data set. The next 35 lines of data took another 6 hours of iterative coding and prompted the visualizing process to help describe relationships between categories. As the model grew, the paper model moved to the wall, where it became increasingly modular and expansive (Figure 2). Although the initial, undocumented analysis of a year earlier assisted in a faster identification of the data categories, frequent adjustments were made to the model structure until the statements of Participant 5 were coded. After this, the model was able to house all further examples of project selection elements, and it seemed that theoretical saturation had been reached even though more practical examples of the different process phases were still being added. The subsequent structured online survey was used to identify more practical examples from a larger (*n=182*) participant group, but only five more examples were added, and no structural changes were required from the model.

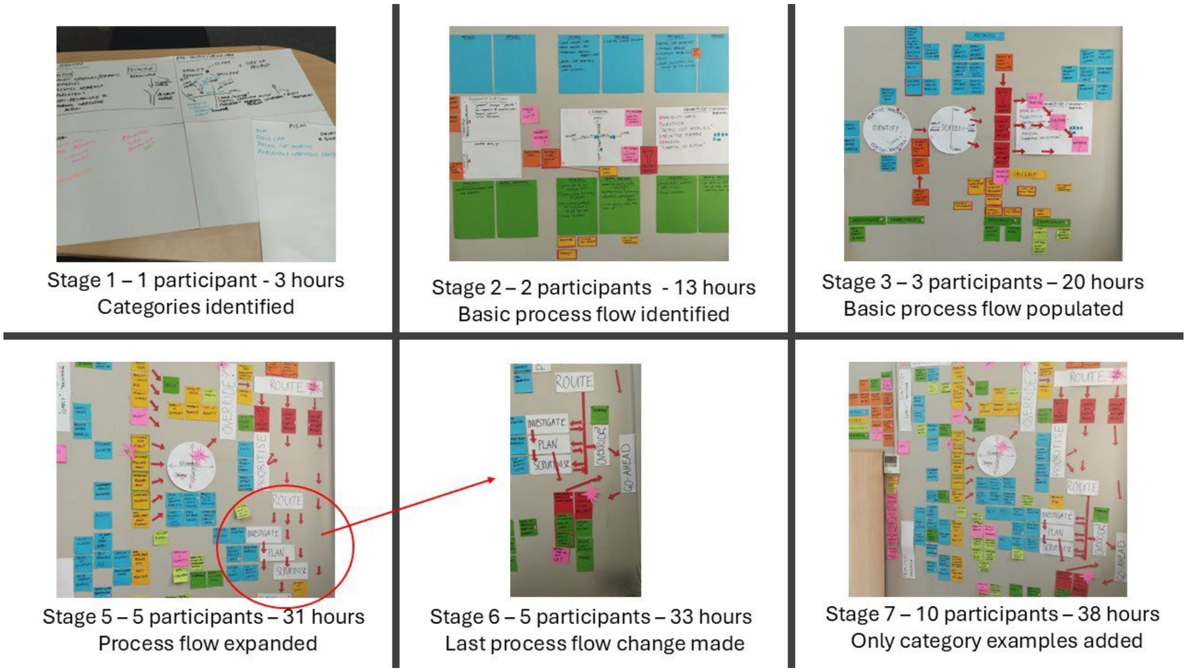
The final model thus displays both the first-order codes (most of the smaller coloured blocks), as well as the second-order codes (the white and red elements) and together create the "anatomy" or "photograph" (Gioia et al., 2013, p. 22), as well as some of the "physiology" or "movie" of the data set. Digitisation, which included streamlining and organising the model, as well as a comprehensive discussion of the model allowed for the presentation of the theoretical model to be described in full in a dedicated publication.

The time study data collected during the process shown in Figure 2 shows how the coding, measured in minutes per line of data, was slow at first, but became more efficient as the developing theoretical model was able to quickly absorb and process the data points toward the end of the analysis. The analysis started with the data of the most experienced practitioner (P2R) who had observed, executed, and lived with project selection decisions for nearly 40 years.

Starting with the richest and most diverse data set, which covered project selection at various levels and in various contexts, played a role in developing a basic and hypothetical model structure (second-order coding) in parallel with the first-order coding. The meticulous and transparent process ensured that first-order coding still stayed true to the original data set. Figure 3 shows how much time was spent on getting an initial grasp on the shape of the data (Participant 2R.1) before a basic structure (the model categories) was in place to be further refined as coding proceeded, still consuming a lot of analytical power (Participant 2R.2 and Participant1G).

Figure 2

Stages of Model Development

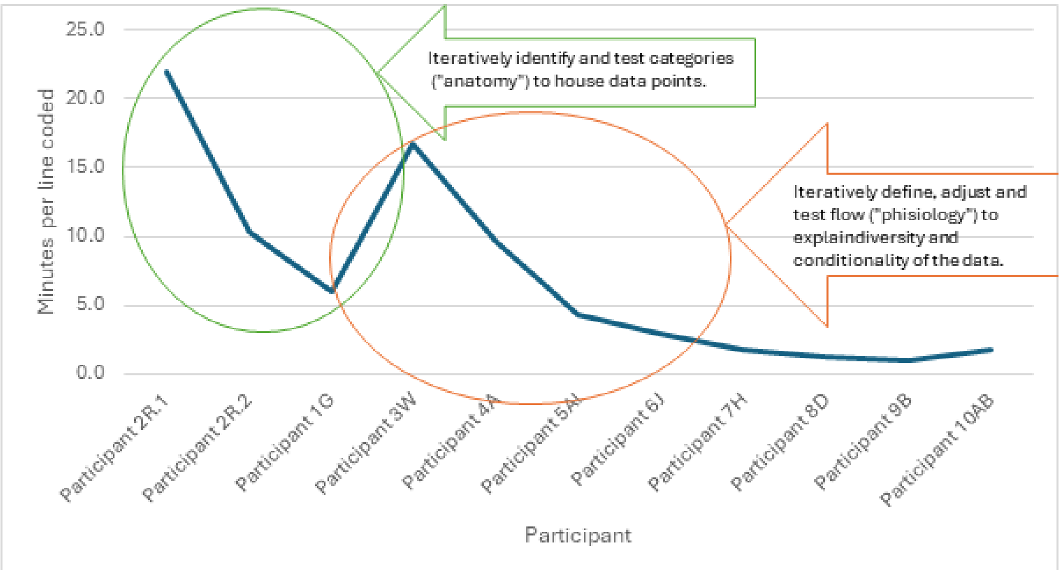


Although the expanding categories ("the photograph") were able to house the data points, a deeper, more complex sequence or flow or physiology ("the movie") needed to be created as the examples, although of the same data category, such as method or criteria, seemed to be used at different stages and under different circumstances. This meant that a more complex structure was needed to house the data points in the right stage and under the right conditions and could not simply be listed together under one category. This complexity was identified after coding around 80 lines of data, coming from the diverse situations observed over a cumulative 100 years of project selection observed by Participants 2R, 1G and 3W.

Further physiological adjustments were made during the analysis of the next 86 lines of data contributed by Participants 3W, 4A and 5Al as the deeper structure was iteratively tested and refined. After this, the remaining 156 lines of data could be slotted into the proposed model with relative ease, meaning that they served to test, confirm and enrich the model.

Figure 3

Time Study Data of the Coding Process



The rate at which coding takes place, in parallel with theory development, is thus proposed as a possible metric with which to track theoretical saturation. The definition of theoretical saturation is, however, the benchmark to determine whether theoretical saturation has been reached: when new data does not contribute any new insights (Hallberg, 2006; Urquhart, 2019). Although the rate at which the proposed theoretical model can absorb new data points can function as an indicator of theoretical saturation, it should be used in conjunction with a tracking measure that monitors the rate at which new contributions are added. Figure 4 shows how the number of novel contributions made by participants decreased as the analysis progressed with participants mostly confirming earlier contributions.

Figures 3 and 4 show how *theoretical* saturation was reached after the analysis of 186 data points contributed by Participants 1 to 5. However, the theoretical model was still being tested and enriched by novel examples of the different categories given. Although no further adjustments were made to the theoretical structure of the model, the novel examples still contributed explanations and challenges to the theoretical model.

To cast a wider net with which to collect examples for, and thus theoretical challenges to, the theoretical model, an online structured survey was selected to be a more efficient tool. The survey tested the prevalence of the already identified model elements to screen for validity, and asked respondents to contribute more examples. The 182 survey respondents contributed another five examples and confirmed the prevalence of all the existing elements.

Figure 4

Number of Novel Contributions Made as the Analysis Progressed

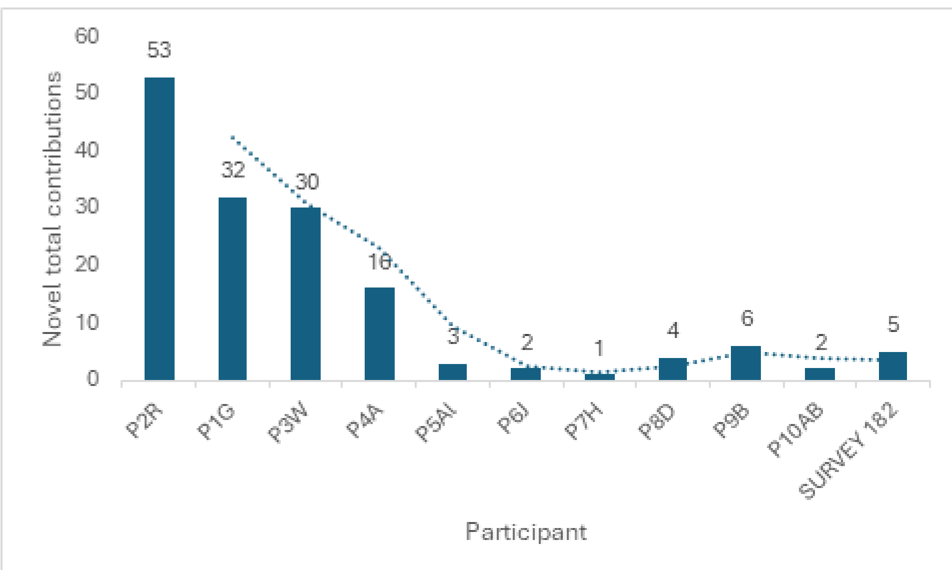
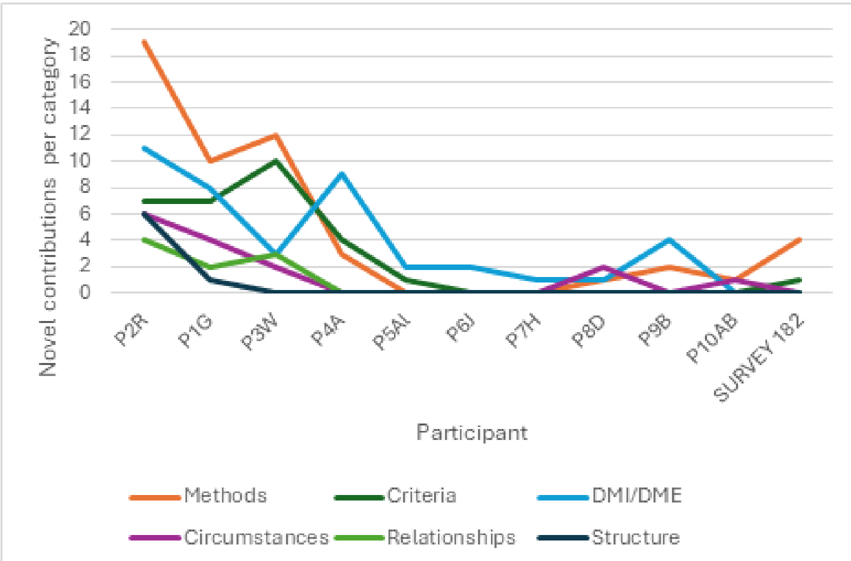


Figure 5 shows a breakdown of the individual contributions, confirming that *theoretical* saturation, measured by the structure ("anatomy") and relationships ("physiology") data sets, was indeed reached after Participant 5. The rest of the contributions were examples of the different categories within the model, e.g. different project selection criteria, or different project selection methods observed, etc. These were thus not theoretical contributions, but rather examples of the theoretical elements.

Figure 5

Breakdown of Novel Contributions Made During the Course of the Analysis



Conclusion

This paper thus proposes and demonstrates an alternative approach to ensuring theoretical saturation, deviating from the blanket rule of thumb sample size approach often adopted in qualitative and GT research. It endorses and encourages a systematic, rigorous data analysis phase, documented and tracked by the researchers for the sake of quality assurance of the data coding process to be rewarded by the gradual emergence of the theoretical contribution.

The paper proposes the following metrics with which to track theoretical saturation:

- Minutes per data point analyzed AND
- Number of novel contributions made per participant data set AND
- Number of theoretical vs non-theoretical contributions made per participant data set.

The ultimate test of theoretical saturation remains the definition of no new theoretical contributions are made.

As superficial or lazy coding could easily lead to a false deduction that theoretical saturation has been reached, this paper also proposes the meticulous and transparent documenting of the (data) point-by-point coding of the entire data set until theoretical saturation has been reached, tested and confirmed satisfactorily. The visual, color-coded approach demonstrated in Figures 1 and 2 makes this proposed rigorous process more accessible to both researchers and potential critics, but is not essential, as the software used to capture the data analysis trail should provide the lasting proof of the rigor of the process.

Classic GT is a powerful research methodology with which to explore real-life phenomena but is easily discredited by its critics due to its open-minded approach that depends on the rigor and objectivity of the researcher. The metrics proposed and demonstrated by this paper can assist GT researchers in ensuring rigor that can easily be proven. It is postulated here that the use of the proposed metrics and approaches demonstrated in this paper will do much to support the credibility of GT research in general.

References

Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. *The SAGE handbook of interview research: The complexity of the craft, 2*, 347-365.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15-31.

Goulding, C. (2002). Grounded theory: A practical guide for management, business, and market researchers,[e-book]. *Available through: LUSEM Library website http://www.lusem. lu. se/library*.

Hallberg, L. R. (2006). The "core category" of grounded theory: Making constant comparisons. *International Journal of Qualitative Studies on Health and Well-being, 1*(3), 141-148.

Kenny, M., & Fourie, R. (2015). Contrasting classic, Straussian, and constructivist grounded theory: Methodological and philosophical conflicts. *The Qualitative Report, 20*(8), 1270-1289.

Thomson, S. B. (2011). Sample size and grounded theory.

Urquhart, C. (2019). Grounded theory's best kept secret: The ability to build theory. *The SAGE handbook of current developments in grounded theory*, 89-106.

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory'back into grounded theory: guidelines for grounded theory studies in information systems. *Information systems journal, 20*(4), 357-381.

Vander Linden, K. L., & Palmieri, P. A. (2023). Developing a classic grounded theory research study protocol: A primer for doctoral students and novice researchers. *Grounded Theory Review, 22*(01), 23-40.

Vander Linden, K. L., & Palmieri, P. A. (2021). Criteria for Assessing a Classic Grounded Theory Study: A Brief Methodological Review with Minimum Reporting Recommendations. *Grounded Theory Review, 20*(2).

Welch, C., & Piekkari, R. (2017). How should we (not) judge the 'quality'of qualitative research? A re-assessment of current evaluative criteria in International Business. *Journal of World Business, 52*(5), 714-725.

Declaration of Conflicting Interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

For all inquiries, contact ilse.doyer@up.ac.za;