RESEARCH · QUALITATIVE RESEARCH

Thematic Analysis: A Step-by-Step Method

A step-by-step guide to thematic analysis, covering what a theme is, the main types of thematic analysis, the six phases from familiarisation to writing, inductive and deductive coding, quality checks, common mistakes, and a worked commercial example.

ModQu research article · 2,489 words · 23 references (APA 7th edition)

Thematic analysis is a method for identifying, analysing and reporting patterns of meaning, called **themes**, across a set of qualitative data such as interview transcripts, focus group discussions, open survey responses or documents. It is the most widely used approach to qualitative analysis, partly because it is flexible: it does not depend on a particular theory of knowledge, and it can be applied to many kinds of question and data. Braun and Clarke (2006) set out a clear six-phase method that made thematic analysis accessible to researchers across many fields, and their paper has become one of the most cited in the social sciences. This article explains what thematic analysis is, describes its main forms, and walks through each phase with practical guidance.

What is a theme?

A theme captures something important about the data in relation to the research question and represents a pattern of shared meaning across the dataset. It is not simply a topic. "Price" is a topic; "price as a signal of trustworthiness" is a theme, because it says something about what participants mean. Braun and Clarke (2019) emphasise that themes are built by the analyst through interpretation; they do not simply "emerge" from the data as if they were waiting to be found.

A theme also need not be the most frequently mentioned issue. A pattern that appears in a minority of accounts can be important if it helps answer the research question. What matters is its relevance and the coherence of the meaning it captures, not a count of mentions.

Three types of thematic analysis

The name "thematic analysis" covers several approaches that share some techniques but differ in assumptions. Braun and Clarke (2021) distinguished three broad types.

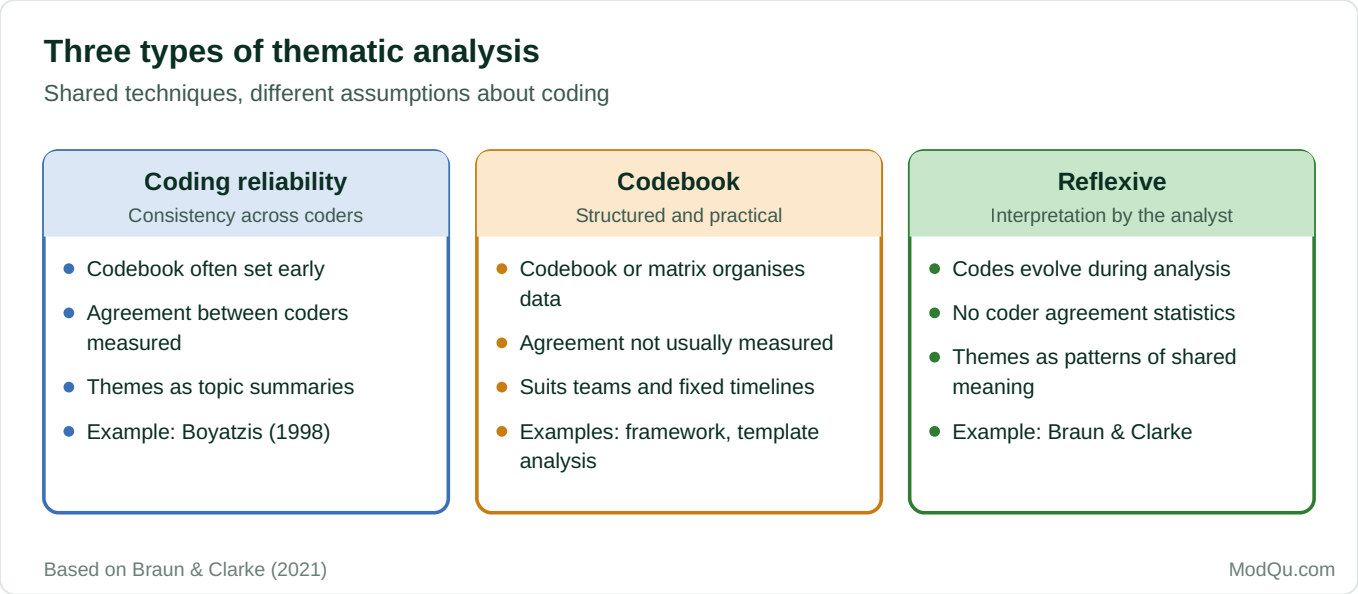


Figure 2. Three broad types of thematic analysis. Based on Braun and Clarke (2021); diagram by ModQu.

Coding reliability approaches, associated with Boyatzis (1998), treat coding as a process that should be consistent across coders. Researchers develop a codebook, often in advance, apply it to the data, and measure agreement between coders. Themes are often defined early and treated as summaries of topics.

Codebook approaches, such as the **framework method** (Ritchie & Spencer, 1994; Gale et al., 2013) and **template analysis** (King, 2004), use a structured codebook or matrix to organise data, but without the emphasis on measuring coder agreement. They are common in applied and policy research, where teams need to work quickly and to compare findings across groups.

Reflexive thematic analysis, the approach Braun and Clarke now emphasise, treats coding as an interpretive process that develops through the analyst's engagement with the data. Codes evolve during the analysis, themes are constructed rather than found, and the researcher's perspective is treated as a resource to be reflected on, not a bias to be eliminated. Measuring agreement between coders is seen as inconsistent with this view, because different analysts are expected to produce different but equally valid interpretations (Braun & Clarke, 2019).

Each approach is legitimate. Problems arise when studies mix them without noticing, for example by claiming a reflexive approach while reporting inter-coder agreement statistics, or by claiming that themes "emerged" from a codebook defined in advance. A study should state which approach it uses and follow its logic consistently.

Inductive and deductive coding

Thematic analysis can be **inductive**, building codes and themes from the data with no prior framework, or **deductive**, using existing theory or the research questions to guide coding. Most studies combine the two. Fereday and Muir-Cochrane (2006) described a hybrid approach in which a template of codes drawn from theory is applied first, with new

codes added when the data contain material the template does not capture. In applied research, a deductive starting point drawn from the research brief, such as the stages of a customer journey, combined with inductive coding within each stage, is often efficient.

Analysis can also focus on **semantic** meaning, what participants explicitly say, or **latent** meaning, the underlying ideas, assumptions and concepts that shape what they say (Braun & Clarke, 2006).

The six phases

Braun and Clarke (2006) set out six phases. They are not strictly sequential; analysts move back and forth between them as the analysis develops.

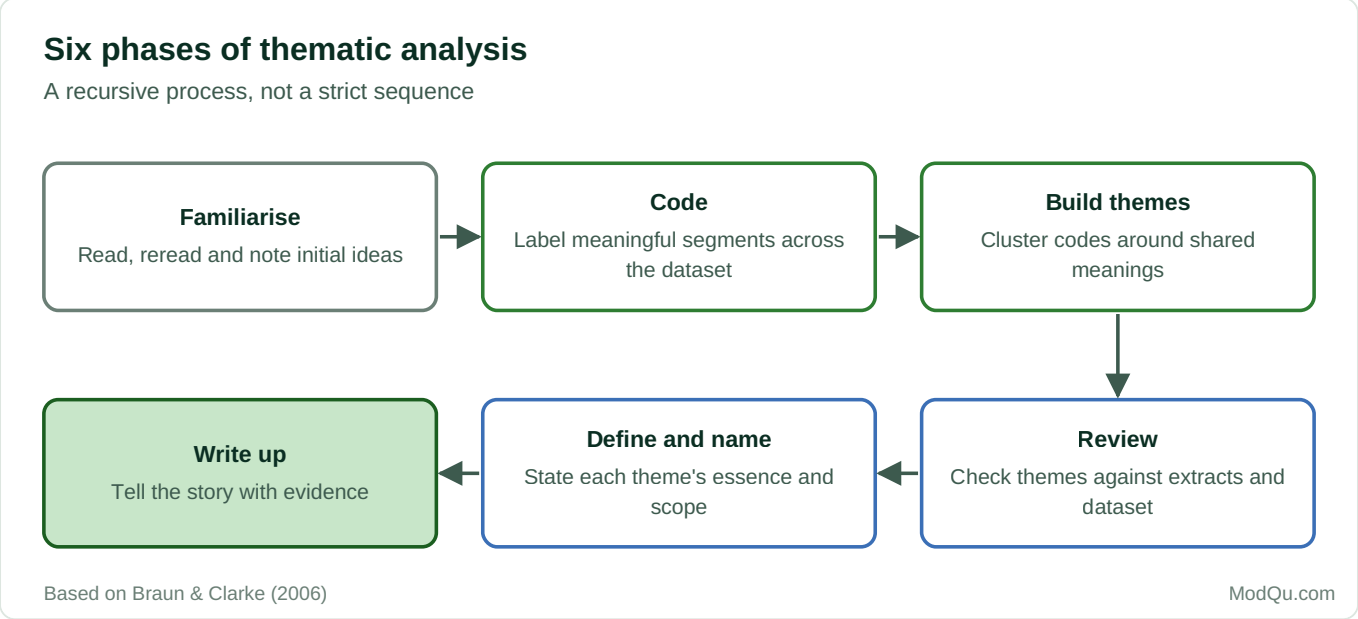


Figure 1. The six phases of thematic analysis. Analysts move back and forth between phases as the analysis develops. Based on Braun and Clarke (2006); diagram by ModQu.

Phase 1: Familiarisation

The analyst reads and rereads the data, listens to recordings and notes initial ideas. Transcribing interviews oneself is a valuable form of familiarisation. The aim is to become immersed in the dataset as a whole before breaking it into pieces. Brief notes about striking points, possible patterns and questions are kept for later phases.

Phase 2: Generating codes

The analyst works systematically through the entire dataset, labelling segments of data that are relevant to the research question. A **code** is a short label that captures the meaning of a segment, such as "fear of hidden costs" or "trust built through small promises kept". Every data item should receive full and equal attention. Saldaña (2016) describes many coding methods, from descriptive codes that summarise topics to process codes and in vivo codes that use participants' own words. A single segment can carry several codes.

Codes should be specific enough to be useful but not so specific that each applies only once. Early in coding, it is better to code generously, capturing anything that might be relevant, because a code that turns out to be unimportant can easily be dropped later, while material that was never coded is easily lost. Keeping a short definition for each code, and an example extract, helps the analyst apply codes consistently and makes the process traceable. As coding proceeds, some codes will be merged, split or renamed; recording these changes is part of the audit trail.

Phase 3: Generating initial themes

The analyst groups codes into candidate themes, looking for clusters of codes that share a central idea. Visual tools, such as tables, mind maps or sorting codes on cards, help. Some codes become themes, some become subthemes, and some are set aside. At this stage, themes are provisional.

Phase 4: Reviewing themes

Candidate themes are checked against the coded extracts and then against the full dataset. Does each theme have a clear central organising concept? Is there enough data to support it? Are the boundaries between themes clear, or do they overlap? Some themes will be merged, some split and some discarded. Braun and Clarke (2006) suggested checking that the data within each theme cohere meaningfully, while there are clear distinctions between themes.

Phase 5: Defining and naming themes

Each theme is defined in a few sentences that capture its essence and scope, and given a concise, informative name. A theme name like "Communication" says little; "Silence read as indifference" tells the reader what the theme is about. The analyst writes a detailed account of each theme and how it relates to the others and to the research question.

Phase 6: Writing up

The report tells the story of the data through the themes, using vivid extracts as evidence and explaining what each shows. Extracts should illustrate analytical points, not replace them. Byrne (2022) provides a detailed worked example of these phases, including the evolution of codes and themes, which is useful for researchers new to the method.

Data displays and matrices

Many analysts find it useful to organise coded data in a **matrix**, with participants in rows and themes or codes in columns, and summaries of what each participant said in the cells. This is the core of the framework method (Gale et al., 2013), but it can support any form of thematic analysis. Matrices make it easy to see which themes are common across groups, which are specific to particular types of participant, and where individuals depart from the pattern. Miles et al. (2014) describe a wide range of such displays and argue that well-designed displays are an essential tool for drawing and checking conclusions.

Quality in thematic analysis

The quality of thematic analysis can be judged against the general criteria of trustworthiness for qualitative research (Lincoln & Guba, 1985). Nowell et al. (2017) showed how each phase can be documented to meet these criteria: keeping raw data and coding records, writing reflexive notes, discussing codes with colleagues, keeping an audit trail of how themes developed, and describing the context in enough detail for readers to judge transferability.

In coding reliability approaches, agreement between independent coders is measured. O'Connor and Joffe (2020) reviewed the debate and argued that inter-coder reliability can be useful in some contexts, particularly when coding frames will be applied by teams or when findings need to persuade audiences used to quantitative standards, but that it is not appropriate for every study. Campbell et al. (2013) showed that agreement depends heavily on how the text is divided into units for coding, and proposed procedures for achieving acceptable agreement when it is required.

Braun and Clarke (2021) set out a checklist for quality in reflexive thematic analysis. It includes questions such as: Are themes clearly defined as patterns of shared meaning, not topic summaries? Is the analysis interpretive, going beyond description? Is the researcher's role acknowledged? Do the claims match the data? Malterud (2001) similarly stressed that qualitative analysis should be transparent about the researcher's position and the steps taken from data to conclusions.

Thematic analysis and content analysis

Thematic analysis is often confused with **qualitative content analysis**. Vaismoradi et al. (2013) noted that both identify patterns in text, but that content analysis tends to focus on categorising and sometimes counting manifest content, while thematic analysis gives more weight to interpretation and to the meaning of patterns. Hsieh and Shannon (2005) distinguished three forms of content analysis: conventional (categories derived from data), directed (categories from theory) and summative (counting words or content, followed by interpretation). Both methods are valid; the choice depends on whether the research needs descriptive categories, frequencies or interpretive themes.

Working in teams

Commercial and applied projects are often analysed by teams under time pressure. Teamwork can strengthen an analysis, because different analysts notice different things, but it needs coordination. A common approach is for two or three analysts to code the same few transcripts independently, meet to compare and discuss their codes, and agree a shared set

of codes and definitions before dividing the remaining data. Regular meetings during coding allow new codes to be shared and definitions refined. In a reflexive approach, the aim of these discussions is richer interpretation, not agreement statistics; in a codebook or coding reliability approach, the aim is consistent application of the codebook (Braun & Clarke, 2022).

Teams should also agree who has final responsibility for the themes. Analysis by committee can produce bland, safe themes that everyone can accept, rather than sharp interpretations that capture what is distinctive in the data.

Using software

Qualitative data analysis software, such as NVivo, ATLAS.ti or MAXQDA, helps manage large datasets: it stores transcripts, allows codes to be applied and retrieved, and supports matrices and searches. Spreadsheets and word processors can serve the same purposes for smaller studies. Software does not perform the analysis. It makes it easier to keep track of codes and extracts, and to check how a theme is distributed across participants, but the interpretation remains the analyst's work (Saldaña, 2016). Newer tools that use language models to suggest codes or summarise transcripts can speed up early phases, but their suggestions need to be checked against the data, and the analyst remains responsible for the themes and the claims made about them.

Reporting

A report of thematic analysis should state which type of thematic analysis was used and why; describe the data and how they were collected; explain the coding process, including whether coding was inductive, deductive or both; describe how themes were developed and reviewed; and present each theme with a definition, an interpretation and supporting extracts. For interview and focus group studies, the COREQ checklist provides a useful list of items to include (Tong et al., 2007). Braun and Clarke (2022) recommend that reports present a thematic map or table showing the themes and subthemes, so that readers can see the overall structure of the analysis before reading the detail.

Common mistakes

Several mistakes recur in published and commercial thematic analyses:

- **Themes as topics.** Reporting "themes" that are simply the questions asked in the interview guide, such as "views on price" or "views on service".
- **Too many themes.** Twenty themes usually means that coding has not been taken far enough; most analyses produce between two and six main themes, some with subthemes.
- **Description without analysis.** Presenting a string of quotations with little interpretation of what they mean.
- **Counting as evidence.** Treating the number of participants who mentioned something as the measure of its importance.
- **Unclear method.** Describing the analysis only as "thematic analysis was conducted" without saying which approach was followed or how.

Clarke and Braun (2017) emphasise that the method's flexibility is a strength only if researchers make their choices explicit.

A worked commercial example

Consider a hypothetical study of why small business owners switch accounting software. The research team conducts 20 interviews and uses reflexive thematic analysis. In familiarisation, they note that many owners describe the decision emotionally, not technically. In coding, they produce codes such as "dread at month end", "feeling stupid with the software", "accountant's recommendation", "price rise as last straw" and "fear of losing data".

In generating themes, they cluster these into three candidate themes: the software as a source of anxiety, the accountant as a trusted decision maker, and switching as a risk to be managed. Reviewing the themes against the data, they find that "price rise as last straw" fits poorly with any theme and appears to act as a trigger rather than a reason, so they create a fourth theme on triggers. In defining themes, they name them "Software that makes me feel incompetent", "The accountant decides", "Switching feels dangerous" and "The final straw". The report shows that product improvements aimed at technical features will matter less than reducing anxiety and working through accountants, and it presents extracts that show each theme in participants' words.

Guest et al. (2012) and Kiger and Varpio (2020) offer further practical guidance for applied teams working to fixed timelines, including how to divide coding across team members and how to keep analysis consistent.

Conclusion

Thematic analysis is a flexible and accessible method for finding patterns of meaning in qualitative data. Its value depends on clarity about which form is being used, systematic work through the phases of familiarisation, coding, theme development, review, definition and writing, and themes that interpret the data rather than merely summarise topics. Done well, it turns hours of conversation into a small number of clear, evidence-based insights that can guide understanding and action.

References

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. SAGE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity, 56*(3), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research, 42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80–92. https://doi.org/10.1177/160940690600500107

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology, 13*, Article 117. https://doi.org/10.1186/1471-2288-13-117

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. SAGE. https://doi.org/10.4135/9781483384436

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher, 42*(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

King, N. (2004). Using templates in the thematic analysis of text. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 256–270). SAGE.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet, 358*(9280), 483–488. https://doi.org/10.1016/S0140-6736(01)05627-6

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1). https://doi.org/10.1177/1609406917733847

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406919899220

Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 173–194). Routledge.

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, *19*(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398–405. https://doi.org/10.1111/nhs.12048