A Multi-level Coding Method for Chinese Free Narrative and Its Application in Psychological Research

Authors: Liang Wanying, Yuqing Zhang, Yuqing Zhang

Date: 2026-04-20T19:16:40+00:00

Abstract

Qualitative research occupies an important position in the field of psychology; however, currently popular qualitative research methods find it difficult to simultaneously balance the characteristics of natural flexibility, clear procedures, and rigorous standards. In contrast, a multi-level coding method designed for Chinese free narrative possesses both natural and comprehensive characteristics and is relatively standardized.

This method consists of three stages: partitioning large blocks of text to form major categories, extracting information granules of different sizes to form subcategories, and connecting subcategories. It shares similarities with Western grounded theory and the "Puxue" (philology) of the Chinese Qing Dynasty but maintains essential differences.

This method offers specific value to psychological research, such as supporting mixed analysis and self-verification (falsification), possessing the potential to analyze the unconscious, providing possibilities for discovering unexpected correlations, and improving the precision of the transition from qualitative to quantitative analysis. At the same time, it has limitations, including risks associated with standalone use and high time costs for implementation. These limitations await breakthroughs by improving existing algorithmic tools and determining how each step can respectively rely on algorithmic tools and researcher implementation.

Full Text

Evaluation of a Multi-level Coding Method for Chinese Free Narrative and Its Value for Psychological Research

Center for Adolescent Personality and Health Research, Institute of Psychology, Chinese Academy of Sciences; Department of Psychology, University of Chicago

Abstract

Qualitative research occupies a significant position within the field of psychology; however, currently popular qualitative methodologies often struggle to balance the requirement for rigor with practical flexibility. To address this, a multi-level coding method designed for Chinese free-form narratives has been developed, characterized by its naturalistic and comprehensive qualities while remaining relatively standardized. This method consists of three distinct stages: partitioning large blocks of text to establish primary categories, extracting information granules of varying sizes to form subcategories, and establishing connections between these subcategories. While this approach shares similarities with Western Grounded Theory and the *Puxue* (evidential scholarship) tradition of China' s Qing Dynasty, it maintains fundamental distinctions from both.

This method offers unique value to psychological research by supporting mixed-methods analysis and self-verification (or falsification), uncovering latent analytical meanings, and providing the potential to discover unexpected correlations. Furthermore, it enhances the precision of the transition from qualitative to quantitative data. Despite these advantages, the method carries risks when used in isolation and involves high time costs for implementation. Future developments should focus on refining existing algorithmic tools and determining how each step can be optimally executed through a collaboration between researchers and computational tools. We invite readers to test and critically evaluate this methodology.

Keywords

Free narrative; Qualitative analysis; Data mining

---

1 Introduction

Naturally complete and uniquely insightful, qualitative research occupies a vital position within the field of psychology. It is widely recognized in the academic community that qualitative research offers superior exploratory and interpretive depth compared to quantitative methods, although it is often perceived to have lower formal validity. In psychological studies characterized by high levels of structural rigor and the primary goal of hypothesis testing, qualitative methods

are employed less frequently. Conversely, in exploratory research where problem structures are less defined, qualitative research is exceptionally prevalent.

Among the qualitative research methodologies commonly employed today—such as content analysis and case studies—bibliometric analysis of literature indexed in the CSSCI database reveals that thematic analysis and grounded theory account for nearly half of all qualitative research in psychology. Thematic analysis follows relatively structured principles and procedures, effectively satisfying the need to extract meaningful information from qualitative data [**?**]. In contrast, grounded theory excels at helping researchers abstract complex concepts from disorganized information to construct structured theoretical systems [**?**, **?**, **?**].

However, novice researchers who opt for thematic analysis may miss the opportunity for theory construction, while those choosing grounded theory may produce unreliable frameworks due to a lack of academic depth. For experienced researchers, grounded theory carries the risk of a priori interpretations, where solid theoretical backgrounds lead to the omission of latent, valuable information.

Content analysis has also evolved alongside information technology. While traditional manual content analysis was notoriously time-consuming, artificial intelligence (AI) has significantly increased its efficiency. AI-assisted content analysis serves as an ideal tool for improving efficiency, yet it requires researchers to select appropriate algorithms and formulate logical instructions based on specific objectives. To date, standardized workflows for these AI-assisted paths remain scarce. A methodology that is as naturally fluid as Grounded Theory, as procedurally clear as Thematic Analysis, and as rigorous as Content Analysis would be of significant benefit to the field. This article introduces a systematic yet flexible approach to qualitative inquiry that bridges these traditional methodological boundaries.

2 The Multi-level Coding Method

2.1 Methodological Development

The multi-level coding method was developed sequentially in the fields of psychology and education. Liang et al. initially sought a data collection and analysis method that balanced openness with closure to capture novel findings while integrating with established international scholarship. Drawing on qualitative classics and domestic scholarship, they employed semi-structured interviews where participants narrated freely. This prototype was later upgraded by adding specific regulations regarding the nature and size of "information granules."

To enhance trustworthiness, it is essential to report methodological details [**?**, **?**, **?**]. For clarity, we present a standardized workflow using NVivo software to analyze interview transcripts:

Phase 1: Partitioning Text Blocks. The researcher establishes broad categories. This can be bottom-up (e.g., "Pre-university," "Post-university" based

on transcript trends) or top-down (e.g., using "Self-Efficacy Theory" to create nodes like "Mastery Experiences").

Phase 2: Extracting Information Granules. A line-by-line analysis identifies "reference points" coded as nodes. Sub-categories may include time, place, emotion, cause, or causality. Crucially, at least two sizes of information granules are extracted: - **Small-granule information:** Defined at the level of words or phrases (e.g., specific nouns or particles). - **Medium-granule information:** Defined at the level of sentences or multiple sentences (e.g., a specific explanation of a conflict). - **Large-granule information:** Reference points that may span significant distances across the text or even cross the boundaries of the large categories formed in Phase 1.

Phase 3: Connecting and Conceptualizing. Sub-categories are linked to reflect dynamic or static relationships (coordination, progression, or subordination). This results in a node tree or network that represents the final theoretical concepts.

2.2 Characteristics: Naturalness and Exhaustiveness

This method is "natural," making it suitable for free-narrative texts with low structure. It is also "exhaustive," as the coding manual sets rules for all content —including seemingly insignificant structural particles—regardless of perceived relevance to the theme. This ensures that the data "speaks for itself" before the researcher's subjective interpretation takes over.

While it shares similarities with Grounded Theory (bottom-up logic, researcher as the primary instrument), it is fundamentally different. In Grounded Theory, granules are "meaning-based"and involve immediate cognitive processing. In this multi-level method, small-granule extraction is lexically driven and mandatory, preserving the original material as objective evidence regardless of the researcher's initial focus.

3 Methodological Comparisons

The logic of reducing data to its original meaning finds a historical parallel not in Western methodology, but in the *Puxue* (Plain Learning) tradition of China's Qing Dynasty. *Puxue* scholars countered "empty talk" through rigorous textual criticism, seeking to reconstruct the original meaning of classics by analyzing individual characters, phonology, and semantics [**?**, **?**]. This marked a "linguistic turn" where psychological speculation was replaced by philological rigor.

However, while *Puxue* was limited to canonical scriptures, the current method focuses on naturally generated free-narrative transcripts. While Grounded Theory is "natural"but often lacks "completeness,"and *Puxue* is "complete"but lacks "naturalness,"the proposed multi-level coding method balances both, supporting the transition from minute evidence to grand meaning construction.

4 Value for Psychological Research

4.1 Mixed-Methods Analysis and Self-Verification

The completion of coding allows qualitative results to be derived into quantitative metrics (number of nodes and reference points). Because the method extracts different granularities, it allows for internal validation. For example, if medium-grained data suggests a participant values "Parental Opinion," but small-grained data shows a significantly higher frequency of the word "Spouse," the researcher can identify a discrepancy between conscious reporting and latent priorities.

4.2 Analyzing the Subconscious

Extracting fine-grained information can reflect content mentioned at an unconscious level. In unstructured data, such as childhood narratives, this method helps distinguish between conscious self-presentation and subconscious preoccupations, providing a deeper layer of psychological insight.

4.3 Discovering Unexpected Correlations

The transition from qualitative to quantitative logic enables the discovery of correlations overlooked by traditional thematic analysis. For instance, correlation analysis of small-grained reference points might reveal a link between childhood peer contact and adult partner choice ($r = 0.863$) [**?**]. These quantitative insights can inspire new theoretical directions.

4.4 Precision in Qualitative-to-Quantitative Mapping

Compared to traditional content analysis, this method improves mapping accuracy through a solid chain of evidence. By distinguishing between the frequency of a topic (small-grain) and the sentiment expressed toward it (medium-grain), researchers can avoid crude conclusions and capture the nuances of participant needs and skepticism.

5 Limitations and Future Directions

The multi-level coding method is a tool, not a standalone methodology. It should be used to support established frameworks like thematic analysis or grounded theory. Using it in isolation risks a mechanical deconstruction of data that lacks theoretical cohesion.

Furthermore, the time cost is substantial. Manual coding across three levels can take months of daily labor. Current AI and Large Language Models (LLMs) still struggle with context-dependent pronouns, irony, and the deep emotional connections required for Phase 3. Future research should focus on human-machine collaboration, determining which steps can be automated to overcome the barri-

ers of labor costs while maintaining the interpretive depth essential to psychology.

6 Conclusion

The multi-level coding method integrates the strengths of Grounded Theory, thematic analysis, and content analysis while addressing their limitations. As an indigenous method developed for Chinese linguistic corpora, it embodies the tradition of "no belief without evidence." We invite the academic community to test and refine this approach to enhance the rigor and exploratory power of qualitative psychological research.

Note: Figure translations are in progress. See original paper for figures.