2011 International Symposium on Empirical Software Engineering and Measurement

Recommended Steps for Thematic Synthesis in Software Engineering

Daniela S. Cruzes
Dept. of Computer and Information Science, NTNU
NO-7491 Trondheim, Norway
dcruzes@idi.ntnu.no

Tore Dybå
SINTEF
NO-7465 Trondheim, Norway
tore.dyba@sintef.no

***Abstract*—Thematic analysis is an approach that is often used for identifying, analyzing, and reporting patterns (themes) within data in primary qualitative research. 'Thematic synthesis' draws on the principles of thematic analysis and identifies the recurring themes or issues from multiple studies, interprets and explains these themes, and draws conclusions in systematic reviews. This paper conceptualizes the thematic synthesis approach in software engineering as a scientific inquiry involving five steps that parallel those of primary research. The process and outcome associated with each step are described and illustrated with examples from systematic reviews in software engineering.**

Keywords—research synthesis; secondary research; systematic review; evidence-based and empirical software engineering

I. INTRODUCTION

Research synthesis is a collective term for a family of methods that are used to summarize, integrate, combine, and compare the findings of primary studies on a specific topic or research question [15][16]. These methods embody the idea of making a new whole out of the parts to provide novel concepts and higher-order interpretations, novel explanatory frameworks, an argument, or new or enhanced theories or conclusions. Such syntheses can also identify crucial areas and questions for future studies that have not been addressed adequately with past empirical research. Research synthesis is built upon the observation that no matter how well designed and executed, empirical findings from single studies are limited in the extent to which they may be generalized [12].

A number of different methods have been proposed for the synthesis of qualitative and mixed-methods findings such as the ones we typically find in software engineering (SE) (see [15] for an overview). Many of the these methods have much in common with meta-ethnography, as originally described by Noblit and Hare [37], and used in SE by Dybå and Dingsøyr [21]. This method involves identifying key concepts from primary studies, comparing and translating them into higher-order interpretations. It is these interpretations, the idea of 'going beyond' the content of the primary studies to perform some form of conceptual innovation, that is key to the synthesis [9].

Like meta-ethnography, many of the other synthesis methods are based on approaches used in primary research. Of these, thematic analysis stands out as it represents a range of potential methods for research synthesis that can be used with most, if not all, qualitative methods [8]. Thematic analysis is also one of the most frequently used synthesis methods in SE; in a previous study we found that two-thirds of the systematic reviews in SE that synthesize their primary studies, performed a narrative or thematic analysis [15].

Thematic analysis is a method for identifying, analyzing, and reporting patterns (themes) within data. It minimally organizes and describes the data set in rich detail and frequently interprets various aspects of the research topic. Thematic analysis can be used within different theoretical frameworks, and it can be an essentialist or realist method that reports experience, meanings, and the reality of participants [8]. It can also be a constructionist method, which examines the ways in which events, realities, meanings, experience, and other aspects affect the range of discourses. Thematic analysis has limited interpretative power beyond mere description if it is not used within an existing theoretical framework [8].

Table 1 shows four approaches for thematic analysis in the process of synthesizing evidence from a set of primary studies. Braun and Clarke outline what thematic analysis is, locating it in relation to other qualitative analytic methods that search for themes or patterns, and in relation to different epistemological and ontological positions [8]. They provide guidelines to start thematic analysis, or conduct it in a more deliberate and rigorous way, and consider potential pitfalls in conducting thematic analysis. Boyatzis proposes two main approaches for thematic analysis [6]; a quantitative description of the frequency of themes, and the forming of clusters of themes. Attride-Stirling proposes that thematic analysis can be usefully aided by, and presented as, thematic networks: web-like illustrations that summarize the main themes constituting a piece of text [1].

Recently, Thomas and Harden developed a new approach called thematic synthesis [42], which draws on the principles of thematic analysis and other established methods in primary qualitative research. It identifies the recurring themes or issues in the primary studies, analyzes these themes, and draws conclusions in systematic reviews. The purpose is to develop analytical themes through a descriptive synthesis and find explanations relevant to a particular review question. The method was developed to address specific review questions about need, appropriateness, acceptability of interventions, and effectiveness.

Thomas and Harden's process of thematic synthesis involves three steps, which overlap to some degree: the free line-by-line coding of the findings of primary studies; the organization of these 'free codes' into related areas to construct 'descriptive themes'; and the development of 'analytical themes'. The development of descriptive and analytical themes invokes reciprocal translation and constant comparison. As such, thematic synthesis shares some characteristics with both meta-ethnography and grounded theory. In

978-0-7695-4604-9/11 $26.00 © 2011 IEEE
DOI 10.1109/ESEM.2011.36

TABLE 1. MAIN APPROACHES FOR THEMATIC ANALYSIS.

	Braun and Clark [8]	Boyatzis [5]	Attride-Stirling [1]	Thomas and Harden [42]
Discipline	Psychology	Social Science	Health Improvement	Social Science
Data	Raw data	Raw data	Raw data	Primary studies
Coding	Theory or data driven	Theory or data driven	Use of theoretical coding framework	Line-by-line coding
Themes	Use of thematic maps	Constant comparison, scoring, scaling and clustering	Use of thematic maps	Axial coding
Trustworthiness	Transparency	Reliability	Not Mentioned	Transparency, generalizability, quality and sensitivity analysis

the latter case, primary studies become the unit of analysis instead of original texts and transcripts.

We extend existing approaches with relevant guides and recommendations for SE and conceptualize thematic synthesis in SE as a scientific inquiry consisting of five steps based on the extent literature (Table 1). The process and outcome associated with each step parallel those of primary research and are exemplified with the eight systematic reviews that were categorized as thematic synthesis in a tertiary study [15] of synthesis in SE [2][4][30][31][32][41][43][44]. The fact that none of these reviews refer to a method of synthesis is a strong indication that concrete steps and recommendations are needed. From these reviews, we identified a need for a more systematic and transparent process of the thematic synthesis of the evidence in SE and also for a stronger focus on research synthesis in SE.

Fig. 1 shows an overview of the synthesis process, while Table 2 describe the main steps and checklist items we propose for thematic synthesis in SE. In practice, reviewers will move in an iterative manner among the steps. We have separated them out and presented them sequentially simply to provide a structure to the recommendations. We also assume that the steps of planning, identification, and selection of relevant primary studies have already been performed.

II. EXTRACT DATA

A key part of systematic reviews is data extraction, in which essential text and data from the primary studies are obtained in an explicit and consistent way according to a defined extraction strategy. However, before starting the extraction, we recommend reading the entire set of selected papers at least once to get immersed with the data.

Such immersion is important in order to be familiar with the depth and breadth of the evidence. Initial ideas and identification of possible patterns in the data will be shaped during this first reading. Although it is tempting to skip this step, most of the literature on thematic analysis advises strongly against skipping this immersion [8].

Immersion was explicitly stated in half of the thematic syntheses included in the tertiary study. For instance, Staples and Niazi described the following [41]:

> The planned selection process had two parts: an initial selection from the search results of papers that could satisfy the selection criteria, based on a reading of the title and abstract of the papers; followed by a final selection from the initially selected list of papers that satisfy the selection criteria, based on a reading of the entire papers.

After the initial reading of the papers, reviewers can also update the systematic review protocol, which among other things contains the data extraction strategy and synthesis strategy [33][34].

There are several ways of extracting data from papers. We recommend a structured reading technique as proposed by Cruzes et al. to explore the evidence of a systematic review [17]. Using this technique, the reviewer follows a procedure for identifying context information and the findings of the paper. Fig. 2 shows a possible template for this data extraction structure. There is at least one context description associated with each paper (possibly more, if the paper describes data collected from several studies), and there will be at least one finding for each context, probably many. For the context, for example, it is important that the reviewer focuses on information that will help in the understanding and interpretation of the findings of the study.

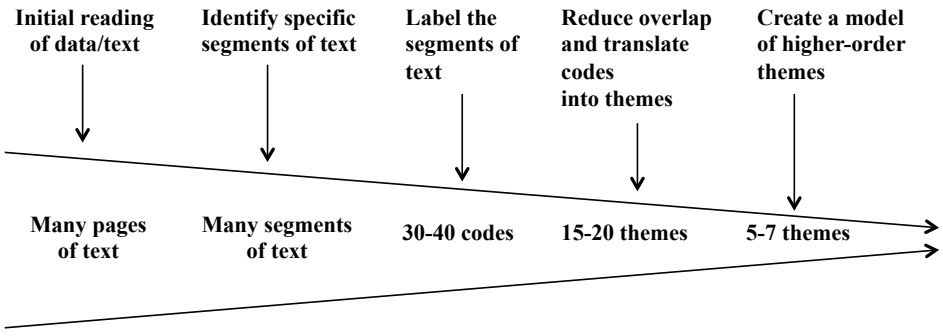

Figure 1. Thematic synthesis process (adapted from Creswell [14]).

TABLE 2. THEMATIC SYNTHESIS STEPS AND CHECKLIST.

Steps	Description	Checklist
Extract data	Extract data from the primary studies, including bibliographical information, aims, context, and results.	1. Have all papers been read carefully to get immersed with the data? 2. Have specific segments of text pertaining to the objectives of the synthesis been identified? 3. Have publication details, context descriptions, and findings been extracted from all papers? 4. Have another researcher checked the extraction?
Code data	Identify and code interesting concepts, categories, findings, and results in a systematic fashion across the entire data set.	5. Have important segments of text like concepts, categories, findings, and results been labeled and coded? 6. Has coding been done across the entire data set on a level that is appropriate for the research questions? 7. Has a list of initial codes with definitions and frequencies been created and checked by another researcher? 8. Have consistency checks or inter-rater reliability checks been performed to establish the credibility of the coding? 9. Are there clear, evident connections between the text and the codes?
Translate codes into themes	Translate codes into themes, sub-themes, and higher order themes.	10. Have themes been created from a thorough, inclusive, and comprehensive review of the codes of all papers? 11. Has overlap between codes been reduced and the remaining codes been collated and translated into themes 12. Have themes been checked against each other and back to the data of the original papers? 13. Are themes internally coherent, consistent, and distinctive?
Create a model of higher-order themes	Explore relationships between themes and create a model of higher-order themes.	14. Have themes been compared across studies, translated into each other, and interpreted into higher-order themes? 15. Have higher-order themes and relationships between themes been checked against the research questions of the synthesis? 16. Are there clear descriptions of the higher-order themes and the relationships between these themes? 17. Has a model been created to show the relationships between the higher-order themes?
Assess the trustworthiness of the synthesis	Assess the trustworthiness of the interpretations leading up to the thematic synthesis.	18. Have the assumptions about, and specific approach to, the thematic synthesis been clearly explicated? 19. Is there a good fit between what is claimed and what the evidence shows? 20. Are the language and concepts used in the synthesis consistent? 21. Are the research questions answered based on the evidence of the thematic synthesis?

For any thematic synthesis, therefore, we recommend the extraction of the following three kinds of data:

- *Publication details* (e.g., authors, year, title, source, abstract, aims);
- *Context descriptions* (e.g., subjects, technologies, industry, settings);
- *Findings* (e.g., results, behaviors, actions, phenomena, events, quotes).

When relevant data are found during the initial reading, it can be useful to highlight them or to transfer key details to a data form, or to code directly into a qualitative analysis tool.

Publication details are usually extracted straightforward, but the aims of the studies can often be unclear and some analytical work may be necessary to identify them.

Context information is more challenging, because some papers lack sufficient details about the design and findings of a study [23]. Sometimes methods are not adequately described, issues of bias and validity might not be addressed, or methods of data collection and analysis, and samples and study settings might not be well described. In some of these cases, data extraction will be hindered.

For the extraction of findings, the key sources in the text of the primary studies are most likely to be found in the sections describing the results, the analysis of results, discussion, and conclusions. Tables and figures are also source of findings. Although not explicitly stated in the text of the paper, relationships visually expressed in figures can also be extracted and translated into textual form. When identifying a statement that could potentially be a finding, the following questions might help [17]:

- Does it state the results of measurements?
- Does it summarize raw data?
- Does it highlight some specific characteristic of the raw data?
- Does it provide additional insights about tables or figures?
- Does it summarize the results of analyses?

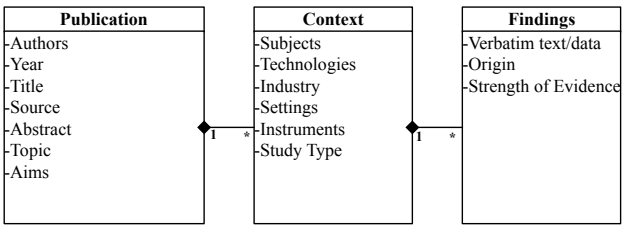

Figure 2. Extraction template.

- Can it be used to answer the research question(s)?
- Does it reflect the main results of the study?

Staples and Niazi, for example, performed the extraction as follows (see Fig. 3) [41]:

> In the data extraction phase, one researcher read every selected paper and extracted information structured according to the data model shown in Fig. ... The data model consists of three types of entities: Publications, which report one or more Studies, which contain one or more Quotes.

Beecham et al. in a systematic literature review of motivation in SE described their extraction process as [2]:

> We used Endnote version 9 (www.endnote.com) to record reference details for each study. How each study answers the research question(s) was recorded on a separate results form: RQ1: What are the characteristics of Software Engineers? RQ2: What (de)motivates Software Engineers to be more (less) productive? RQ3: What are the external signs or outcomes of (de)motivated Software Engineers? RQ4: What aspects of Software Engineering (de)motivate Software Engineers? RQ5: What models of motivation exist in Software Engineering?

Only two of the thematic syntheses included in the tertiary study did not follow the structure of publication, context, and findings to extract data. Whenever feasible, data should preferably be extracted independently by two or more researchers. The extractions should then be compared and disagreements resolved either by consensus or arbitration by an additional independent researcher [33][34]. However, in SE it is still often the case that articles lack sufficient details about the design and findings of a study, which will hinder data extraction. In these cases, we recommend to extract data in consensus meetings [23]. Uncertainties about any primary sources for which agreement cannot be reached should be investigated as part of sensitivity analyses or evaluations of the trustworthiness of the synthesis.

III. Code Data

Codes are descriptive labels that are applied to segments of text from each study. Coding is the process of examining and organizing the data contained in each study of the systematic review. It is more than applying a label, coding requires a clear sense of the context in which findings are made. It involves identifying one or more passages in the text that exemplify the same theoretical or descriptive idea [25].

Coding is thus a method that enables the researcher to organize and group similar data into categories because they share some characteristic – the beginning of themes. Classification reasoning plus tacit and intuitive senses are used to determine which data 'look alike' and 'feel alike' when grouping them together [39]. Rarely, however, will anyone get coding right the first time. As the researcher code, codes and categories get more refined.

Writers in qualitative research use a variety of terms to talk about codes such as indices, categories, labels, and concepts:

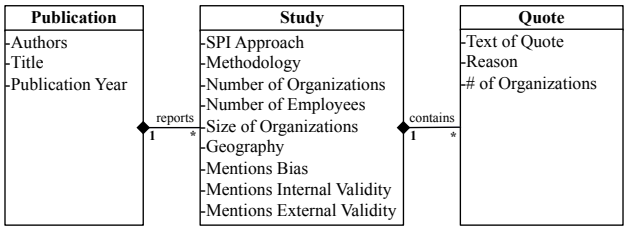

Figure 3. Figure from Staples and Niazi [41].

> Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst, and refer to 'the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon'[6][8].

> Codes are tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study. Codes usually are attached to "chunks" of varying size –words, phrases, sentences, or whole paragraphs, connected or unconnected to a specific setting. They can take the form of a straightforward category label or a more complex one (e.g. a metaphor) [36].

> Coding means naming segments of data with a label that simultaneously categorizes, summarizes, and accounts for each piece of data. Coding is the first step in moving beyond concrete statements in the data to making analytic interpretations [11].

Because thematic data analysis techniques have often adopted the constant comparative method, it has been compared to content analysis and grounded theory. While content analysis consists of comparing and sorting, its aim is to quantify (according to predetermined categories) content in a systematic and reliable manner [10]. It differs from thematic analysis because it, in large part, establishes significance by frequency of themes [25]. On the other hand, some (e.g. [11][13]) distinguish grounded theory from thematic analysis by the unit of text coded; in grounded theory, for example, text can be coded incident by incident or line by line. In contrast, thematic analysis does not specify a coding technique for associating a particular length of text to code [25]. Also, the line-by-line coding recommended by Thomas and Harden, is unlikely to be practical for large numbers of studies.

In systematic reviews in SE it is more appropriate to work with chunks of data as extracted in the previous step. Depending upon the research questions the researcher will focus on the extracted findings alone, or on some combination of findings and context information. In general, the process of coding can be performed according to three approaches:

- *Deductive or A Priori Approach* [36]: The deductive approach starts with creating a provisional 'start list' of codes. This list comes from the theories, research questions, hypotheses, problem areas, and/or key variables that the researcher brings to the study. A start list will typically have from a dozen or so up to 30-40 codes; a number that can be kept well in the analyst's

short-term memory without constant reference to the full list – if the list has a clear structure and rationale. In this approach, the initial step defines a structure of initial codes before the actual coding of the data. These preliminary codes can help researchers integrate concepts already well known in the extant literature. Even though such a 'start list' allows new inquiries to benefit from and build on previous insights in the field, great care must be taken to avoid forcing data into these categories because a code exists for them.

- *Inductive or Grounded Theory Approach* [13]: For grounded theorists, the recommended approach to develop a set of codes is purely inductive. This approach limits researchers from erroneously 'forcing' a preconceived result. Data are reviewed line by line in detail and as a concept becomes apparent, a code is assigned. Upon further review of data, the analyst continues to assign codes that reflect the emerging concepts, highlighting and coding lines, paragraphs, or segments that illustrate the chosen concept. As more data are reviewed, the specifications of codes are developed and refined to fit the data. To ascertain whether a code is appropriately assigned, the analyst compares text segments to segments that have been previously assigned the same code and decides whether they reflect the same concept. Using this 'constant comparison' method, the researcher refines the dimensions of existing codes and identifies new codes.
- *Integrated Approach* [5][35]: The integrated approach employs both inductive (ground-up) development of codes as well as a deductive organizing framework for code types (start list). It is a partway between the deductive and inductive approaches; that of creating a general accounting scheme for codes that is not content specific, but points to general domains in which codes can be developed inductively. Such schemes help the researcher think about categories in which codes will be developed.

The integrated approach is the most relevant in systematic reviews as they tend to be done on the basis of the theoretical interests guiding the research questions of the review. The reviewer approaches the data with specific questions in mind that he/she wishes to code according to. On the other hand, the reviewer also has to relate to the concepts that the authors of the primary studies have organized their findings around. Whether codes are pre-specified or developed along the way, clear operational definitions are indispensable, allowing them to be applied consistently by a single researcher over time or by multiple researchers concurrently.

Any particular study, of course, may focus on only a few of the categories. There are various methods of coding that can be used [13][35][36]; Saldaña [39] describes at least 30 different methods. However, four coding methods are particularly helpful in generating categories, themes, and theory in SE research:

1) Conceptual codes that identify key concept and essential dimensions of these concepts,
2) Relationship codes that identify links between conceptual codes,
3) Subject codes, which identify subjects' perceptions,
4) Context characteristic codes.

As coding is a critical step in the thematic synthesis process, it is crucial that it is completed with rigor and attention to detail. It is important to give equal attention to all papers, identifying interesting aspects in the data that may form the basis for the next steps.

We recommend this step to be performed by at least two researchers who, as in the extraction of the data, will validate the codes. The codes should have explicit boundaries (definitions) so that they are not interchangeable or redundant, and they should also be limited in scope and focus explicitly on the object of analysis, in order to avoid coding every single sentence in the original text. Qualitative software tools can be very helpful in this stage. Nevertheless, some typical problems may arise while coding and should be taken into consideration:

- Coding at a too general a level.
- Identifying what one wants to see and not what the text is saying.
- Coding out of context

Staples and Niazi described this step as follows [41]:

> At the end of the Data Extraction we had extracted 198 Quote objects, each containing a Text of Quote attribute. Two researchers independently reviewed these in order to derive a list of categories to classify these motivations. Using these two lists as a starting point, the two researchers jointly agreed on a list of 22 categories. The two researchers independently classified every Quote object into these categories. Although there was not initially a good level of inter-rater agreement, differences in opinion were discussed in a joint meeting, in some cases with a third researcher arbitrating. The final categories became the Reason attribute for each Quote objects.

Beecham et al. identified the following text segment in a primary study [2], Frangos [26], and labeled it as a motivator: "Recognition based on objective criteria":

> To support this observation, one simply has to take notice of the turnover rates. Upon conclusion of each appraisal period, company personnel turnover rates miraculously peak [2]. This indicates that engineers seek recognition based on objective criteria and not based on management intuition or personal preferences [2].

Subsequently, this part of the text was coded as "Recognition" (for a high quality, good job done based on objective criteria). Eleven other papers had findings on this motivator as well.

IV. Translate Codes into Themes

The definition and analytic function of a 'theme' varies among writers on the subject, and the term 'theme' is often

used interchangeably with other words such as ‘category’, ‘domain’, ‘phrase’, ‘unit of analysis’, and others. Some helpful definitions of theme are as follows:

> A theme is an outcome of coding, categorization, and analytic reflection, not something that is, in itself coded [39].
>
> A theme at a minimum describes and organizes possible observations or at the maximum interprets aspects of the phenomenon. A theme may be identified at the manifest level (directly observable in the information) or at the latent level (underlying the phenomenon) [6].
>
> A theme is an abstract entity that brings meaning and identity to a recurrent experience and its variant manifestations. As such, a theme captures and unifies the nature or basis of the experience into a meaningful whole [19].

Themes pull together a lot of material into more meaningful and parsimonious units; it is a way of grouping the initial codes into a smaller number of sets (15-20 themes), as shown in Fig. 1. Themes reduce large amounts of codes into a smaller number of analytic units, and help the researcher elaborate a cognitive map; an evolving more integrated schema for understanding local incidents and interactions.

Essentially the process of translating codes into themes is to start to consider how different codes may combine to form an overarching theme. As shown in Fig. 4, as the researcher starts to distantiate from the text, the level of abstraction increases and also the generalizability in the definition of the themes. This is not a single step process; as codes are analyzed, some of the first cycle codes may later be subsumed by other codes, relabeled, or dropped all together. As one progresses in the translation to themes, there may be some rearrangement and reclassification of coded data into different and even new codes. The end of the process is when the researcher saturates the possible themes emerging from the data.

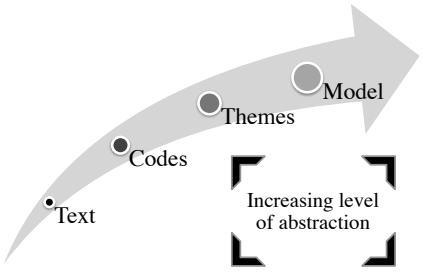

Figure 4. Levels of interpretation in thematic synthesis.

Saldaña describes many different methods for this step [39]: pattern, elaborative or longitudinal coding, and a more grounded theory approach with focused coding [11] and axial and/or theoretical/selective coding [13].

We also recommend the use of visual representations to help sorting the different codes into themes (see for example Fig. 5). The process of deriving themes from textual data and illustrating these with some representational tool is well established in qualitative research. Some tools such as thematic networks, tables, tree-maps, or mind-maps can be used to start organizing the codes and translating them into themes [1][8][40][38].

In the thematic syntheses in SE, this step was usually not described in detail. The authors were usually not explicit about the process of synthesizing the codes and translating them into themes. One exception is Beecham et al., who identified 9 aspects (themes) of SE that motivates software engineers from 21 motivators (codes) [2]:

> Table … identifies themes based on (de)motivators relating to the Software Engineering activity itself. Factors related to salary or other motivators extraneous to Software Engineering itself have not been included in this analysis. This question is an offshoot of our Research Question.

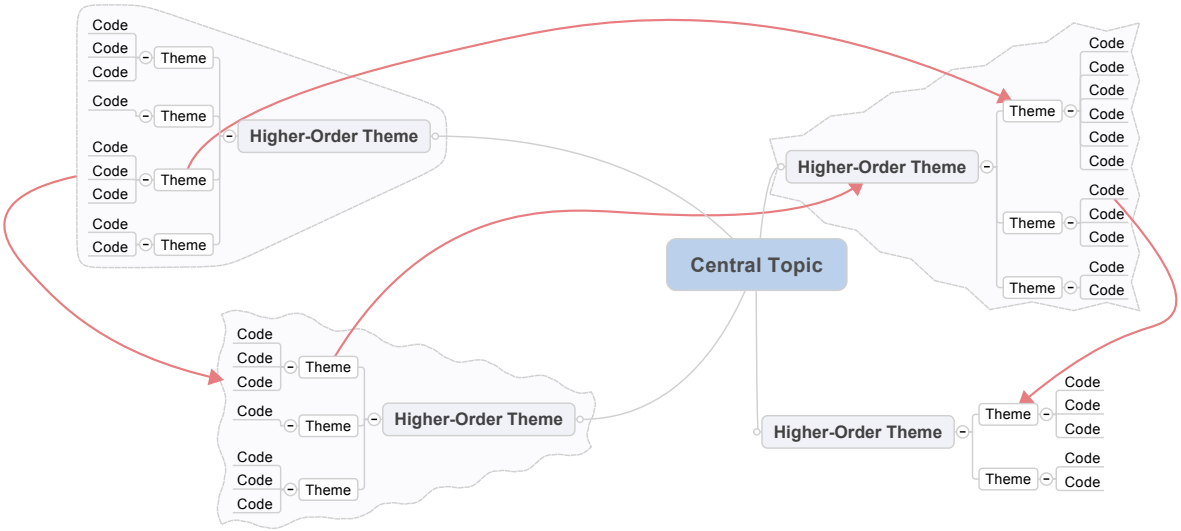

Figure 5. Example of thematic map.

V. Create a Model of Higher-Order Themes

The themes that emerged in the previous steps can now be further explored and interpreted to create a model consisting of higher-order themes and the relationships between them. The aim is to return to the original research questions, and the theoretical interests underpinning them, and address these with arguments grounded in the themes that emerged in the exploration of the texts.

The final product of this step can be a description of higher-order themes, a taxonomy, a model, or a theory. Higher-order themes are recurrent unifying concepts or statements about the subject of inquiry with the purpose of characterizing evidence of individual studies by general insights from the whole set of studies [7]. A taxonomy is a formal system for classifying multifaceted, complex phenomena according to a set of common conceptual domains and dimensions, which is used to increase clarity in defining and comparing complex phenomena [7]. Empirical model building is concerned with constructing practical models useful in describing and coping with real-world situations. Modeling refers to the process of generating a model as a conceptual representation of observed phenomena. A theory is a statement of relationships between units observed or approximated in the empirical world [12][20]. It may be viewed as a system of constructs and variables in which the constructs are related to each other by propositions and the variables are related to each other by hypotheses. Theory is important for understanding potential causal links and confounding variables, for understanding the context within which a phenomenon occurs, and for providing a potential framework for guiding subsequent empirical research.

We recommend the following steps for developing higher-order themes:

1. Review the thematic map from the last step taking each branch in turn, describing its contents with findings and context information and create higher-order themes.
2. Identify connections between the higher-order themes and the underlying evidence and context of this evidence (Arrows in Fig 5).
3. Explore connections with relevant theory and prior research, to recontextualize, define, and further refine the higher-order themes.
4. Create a model, taxonomy, thematic map, or theory of the higher-order themes and their underling evidence.

Some of the primary studies may have reported information about relationships between reported findings and the study context. In these cases, one should compare and contrast the ways in which the relationships have been identified and analyzed across the studies. In other cases little attention may have been paid to these relationships. Then one have to use the data previously extracted from primary studies to look at the relationships between findings and key aspects of the primary studies, and comparing and contrasting these relationships across the studies. This process can be very time consuming but it is critical to the quality of the thematic synthesis as a whole [38].

Exploring the influence of heterogeneity is important at this stage of the synthesis process, because there is considerable heterogeneity in SE studies in terms of findings, methods, contexts, and other unknown sources. Some of the main sources of variability that should be consider when evaluating the robustness of higher-order themes and their relationships are [38]:

- *Variability in outcomes* – the often long causal chains in SE studies may lead to inconsistent results of the same intervention across the studies that the synthesis would ideally seek to address.
- *Variability in study designs* – methodological diversity is common in studies included in systematic reviews in SE. When the main potential sources of variation are known, heterogeneity can be explored by means of subgroup analysis (themes), based for example on theories about how the intervention works, and for which groups.
- *Variability in study populations, interventions, and settings* – the content of complex SE interventions may vary between specific settings or populations. Some of the variability may be intentional as interventions are tailored to context specific needs.

The extent to which it is possible to consider the impact of these sources of variability depends on the availability of relevant information in the primary studies. This might be a challenge, however, since prior research has suggested that there may be a particular problem with inadequate reporting of such information in SE studies [23].

Most authors of systematic reviews in SE describe the themes as found in the previous step without further exploring sources of variability. For example, Khurum and Gorschek showed the frequencies of the primary studies in terms of data extraction category, and the number and percentage of studies from which the corresponding data was extracted [32]. They then used these characteristics to determine the frequencies of the evidence of the studies in an attempt to answer their research questions.

Other authors extrapolate the synthesis by exploring models for assisting the readers in understanding the relationships among themes and underlying conclusions of the systematic review. Beecham et al., for example, used the research questions to discuss the most important themes [2]. They showed a broad picture of what the literature is reporting on motivation in SE and show the characteristics, motivators, and de-motivators of SE. They created a model based on the results of the thematic synthesis (Fig. 6), which shows that the literature is divided as to whether SE forms a distinct occupational group. The majority of the included studies support the idea that these practitioners do form a recognizable group with similar needs. On the other hand, several studies take a contrary view, which lead the authors to conclude that whether or not software engineers form a homogeneous group with similar motivational needs depends on their individual context.

Walia et al., identifyed 14 error types and their meaning in the literature, classified the error types into three higher-order types and showed the results of the classification as a 'requirement error taxonomy' [43]:

> The errors [themes] identified from the software engineering and psychology fields were collected, analyzed for similarities, and grouped into the taxonomy. Errors that had similar characteristics (symptoms), whether from software engineering or from psychology, were grouped into an error class [higher-order theme] to support the identification of related errors when an error is found. An important constraint while grouping the requirement errors was to keep the error classes as orthogonal as possible. For each error class, we first describe the error class. Then we provide a table that lists the specific errors from the literature search that were grouped into that error class. Finally, we give an example of an error from that class, along with a fault likely to be caused by that error.

VI. Assess the Trustworthiness of the Synthesis

Research findings should be as trustworthy as possible and every research study should be evaluated in relation to the methods used. The trustworthiness of a synthesis will depend on both the quality and the quantity of the evidence base it is built on. If primary studies of poor methodological quality are included in the review, trustworthiness of the synthesis can be affected. Trustworthiness will also depend on the methods used, e.g., measures taken to minimize bias, and weighting of studies according to quality [22].

In qualitative research the concepts credibility, confirmability, dependability, and transferability have been used to describe various aspects of trustworthiness [29]:

- *Credibility* deals with the focus of the research and refers to the confidence that can be placed in how well data and processes of analysis address the intended focus. An important issue concerning credibility of a thematic synthesis arises when making a decision about the focus of the study, selection of contexts, participants and the approach to gather primary studies.

  Another critical issue for achieving credibility in thematic syntheses is to extract the most suitable units of data or segments of text. If text segments are too broad, for example, several paragraphs, they can be difficult to manage since they are likely to contain various meanings. Too narrow segments, on the other hand, for example, a single word or a single line of text, may result in fragmentation. In both cases there is a risk of loosing meaning of the text during the coding and abstraction process [28].

  Credibility of a thematic synthesis also deals with how well codes and themes cover data, i.e., no relevant data have been inadvertently or systematically excluded or irrelevant data included. Credibility is also a question of how to judge the similarities within and differences between themes. One approach is to show representative segments of the text. Another is to seek agreement among co-researchers, experts, and participants [28].

- *Confirmability* is concerned with how the extracted data are coded and sorted and whether or not various researchers and experts would agree with the way those data were coded and sorted. Primary study authors' recognition of the findings of the thematic synthesis can also be an aspect of confirmability.

- *Dependability* is concerned with the stability of data, the degree to which data change over time, and alterations made in the researcher's decisions during the synthesis process. Complementary coding methods and establishing an 'audit trail' that will make it possible for an external reviewer to examine the processes whereby data were extracted and coded, and how interpretations and translations into themes were made.

- *Transferability* refers to the extent to which the findings can be transferred to other settings or groups. The authors can give suggestions about transferability, but, ultimately, it is the reader's decision whether or not the findings are transferable to another context. To facilitate transferability, it is valuable to give a clear and distinct description of the selection and characteristics of primary studies, including context and settings, data extraction, and process of synthesis. A rich and vigorous presentation of the findings together with appropriate quotations will also enhance transferability.

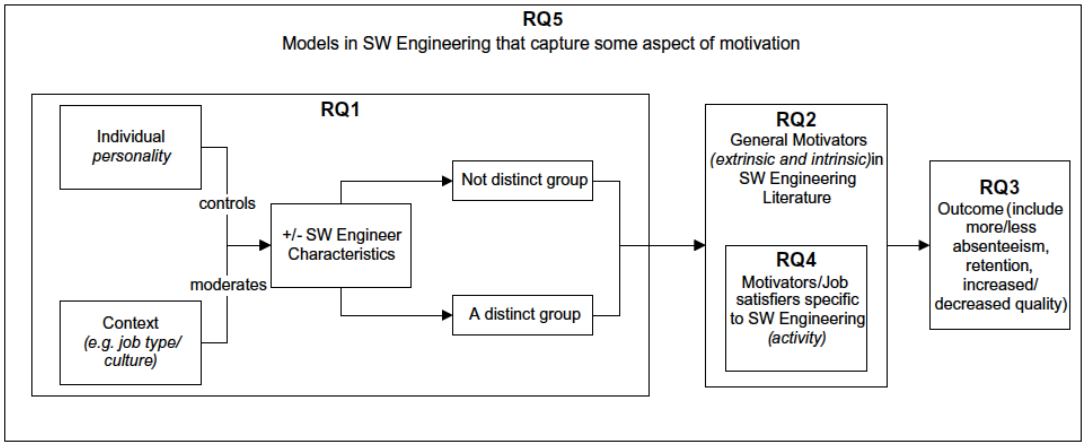

Figure 6. Model of SE motivation from Beecham et al. [2].

Trustworthiness of interpretations made in a thematic synthesis deals with establishing arguments for the most probable interpretations. There is no single correct meaning of research findings, but only the most probable meaning from a particular perspective. Trustworthiness will, therefore, increase if the findings are presented in a way that allows the reader to look for alternative interpretations.

In the tertiary study of synthesis in systematic reviews in SE [15], we found that only 14% of the systematic reviews performed a quality appraisal of the primary studies. None of the thematic syntheses used a quality assessment for evaluating trustworthiness; the papers mainly discussed the bias with the extraction and coding of the results, as for example, Hossain et al. [30]:

> During the data extraction process, we found that several papers lacked sufficient details about the reported projects' contextual factors and the challenges faced and strategies used while using Scrum practices in GDS projects. We synthesized our data by identifying and categorizing the themes from the papers included in this review. Since some of the selected papers do not provide detailed information, there is a possibility that the extraction process may have resulted in some inaccuracies.

Few papers did a deep discussion of the trustworthiness of the synthesis. Beecham et al., however, performed a sensitivity analysis to evaluate possible threats to validity to the synthesis [2]:

> Sensitivity analyses were performed on the studies based on population, location, year and type of study. The sensitivity analyses gave us information on where the data might be biased. … Different countries, areas in Software Engineering and Software Engineer roles have been grouped together in order to identify themes that answer our research questions. However, there is a suggestion in some of the literature that different roles are associated with different motivational needs and characteristics. By grouping all roles together, we may have lost some of this detail. … we are also aware that job titles and responsibilities have changed over the time period covered by the review, e.g. in the mid 1980s the job title 'programmer/analyst' was common, whereas by the early 1990s people were referred to as 'software engineers'. … We may also have lost some of the detail of changes over time by grouping all papers together by theme and ignoring the date of publication in the rage of 1980–2006. … So we have a sample of papers that are more representative of current trends than those in say 1980's (that include 19 papers (or 20%) for the whole decade). When we aggregate our themes the reported frequencies need to be treated with caution.

VII. Discussion and Conclusions

This paper has discussed the rationale for reviewing and synthesizing qualitative research through thematic synthesis. Thematic synthesis has potential to provide a well-organized way of describing large and potentially diverse evidence bases in SE. This paper has also demonstrated how the methods and process of thematic synthesis can be performed in a systematic, comprehensive, critical, and transparent way. Most importantly, this paper draws on a large variety of literature in social science, health, education, psychology, and others to provide a starting point for a more uniform application of thematic synthesis in SE.

The development of synthesis methods for SE brings advantages for the development of relevant theories for SE practice and research. It can also generate new insights from primary research since the synthesis of qualitative research may improve the use of primary research findings, particularly those that relate to rare or infrequent events that are often under-represented in single studies. Another advantage is the enhanced transferability of primary studies. Single qualitative studies can produce in-depth, insightful, and rich descriptions of phenomena, which can be difficult to transfer beyond the original sample. Such studies can, therefore, end up isolated and little used in practice. Synthesis of qualitative research addresses this limitation by drawing single studies together and providing new cumulative knowledge with broader applicability [24].

Several methodological issues remain under debate among those attempting to synthesize qualitative research, such as whether different qualitative research methodologies can be combined, and whether studies included in reviews should be appraised for quality, and if so, with which methods. As more qualitative syntheses are completed, so will such methods become more developed, credible, and transferable.

This paper is also a contribution to current research on how to bring together the findings of qualitative and mixed-methods research to inform policy and practice. Thematic synthesis is by no means the only method available, but by drawing on methods and principles from qualitative primary research, it benefits from the years of methodological development that underpins the research it seeks to synthesize. It is a flexible approach that can be used across a range of epistemologies and research questions to produce an insightful synthesis. However, further work is needed to develop the terminology associated with the synthesis of qualitative research, and also to confirm and refine the recommended steps proposed in this paper.

References

[1] Attride-Stirling, J., "Thematic networks: an analytical tool for qualitative research," Qualitative Research 1(3): 385–405, December 2001.

[2] Beecham, S., Baddoo, N., Hall, T., Robinson, H., and Sharp, H., "Motivation in Software Engineering: A systematic literature review," Inf. Softw. Technol. 50(9-10): 860-878, Aug. 2008.

[3] Bethel, E.C, and Bernard, R.M, "Developments and trends in synthesizing diverse forms of evidence: beyond comparisons between distance education and classroom instruction," Distance Education, 31(3):231-256, 2010.

[4] Bjørnson, F.O. and Dingsøyr, T. "Knowledge management in software engineering: A systematic review of studied concepts," findings and research methods used," Inf. Softw. Technol. 50(11): 1055-1068, Oct. 2008.

[5] Bogdan, R., and Biklen, S.K, Qualitative research for education: An introduction to theory and methods. 2d ed. Boston: Allyn & Bacon, 1982.

[6] Boyatzis, R.E, Transforming qualitative information: thematic analysis and code development, Sage, 1998.

[7] Bradley, E.H, Curry, L.A, and Devers, K.J, "Qualitative Data Analysis for Health Services Research: Developing Taxonomy, Themes, and Theory", Health Services Research, 42(4):1758–1772, August 2007.

[8] Braun V., and Clarke, V., "Using thematic analysis in psychology," Qualitative Research in Psychology, 3(2):77-101. Routledge, 2006.

[9] Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, and M., Pill, R., Using meta ethnography to synthesise qualitative research: a worked example, J Health Serv Res Policy, 7: 209-215, 2002.

[10] Bryman, A., Social Research Methods, 3rd edn. Oxford: Oxford University Press, 2008.

[11] Charmaz, K., Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis, Sage Publications, Beverly Hills, CA, 2006.

[12] Cohen B.P, Developing Sociological Knowledge: Theory and Method, second Ed., Nelson-Hall, Chicago, 1989.

[13] Corbin, J. and Strauss, A., Basics of Qualitative Research (third ed.), Sage, 2007.

[14] Creswell, J.W, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, Prentice Hall, 3rd edition, 2007.

[15] Cruzes, D.S, and Dybå, T., "Research synthesis in software engineering: A tertiary study", Information and Software Technology, 53(5):440-455, May 2011.

[16] Cruzes, D.S, and Dybå, T., "Synthesizing Evidence in Software Engineering Research,'' Proc. of ESEM'10, Bolzano-Bozen, Italy, 16–17 September, pp 1-10, 2010.

[17] Cruzes, D.S, Mendonça, M.G, Basili, V.R, Shull, F., and Jino, M., "Extracting Information from Experimental Software Engineering Papers," Proc. SCCC'07, pp. 105– 114, 2007.

[18] Cruzes, D.S, Mendonça, M.G, Basili, V.R, Shull, F., and Jino, M., "Using Context Distance Measurement to Analyze Results across Studies", Proc. ESEM'07, pp. 235–244, 2007.

[19] DeSantis, L., and Ugarriza, D.N, "The concept of theme as used in qualitative nursing research," Western Journal of Nursing Research, 22(3): 351-372, April, 2000.

[20] Dubin, R., Theory Building, revised edn. New York: Free Press, 1978.

[21] Dybå, T. and Dingsøyr, T., "Empirical Studies of Agile Software Development: A Systematic Review," Information and Software Technology, 50(9-10): 833-859, Aug., 2008.

[22] Dybå, T. and Dingsøyr, T., "Strength of Evidence in Systematic Reviews in Software Engineering," Proc. ESEM'08, Germany, 9-10 Oct., ACM Press, 178-187, 2008.

[23] Dybå, T., Dingsøyr, T., and Hanssen, G.K., "Applying Systematic Reviews to Diverse Study Types: An Experience Report," Proc. ESEM'07, IEEE Press, pp. 225-234, Sept, 2007.

[24] Flemming K., "The synthesis of qualitative research and evidence-based nursing". Evid Based Nurs., 10(3):68–71, 2007.

[25] Floersch, J., Longhofer, J.L, Kranke, D., and Townsend, L., "Integrating Thematic, Grounded Theory and Narrative Analysis: A Case Study of Adolescent Psychotropic Treatment", Qualitative Social Work 9: 407-425, Sept. 2010.

[26] Frangos, S.A, "Motivated humans for reliable software products," Microprocessors and Microsystems, 21(10), 605–610, 1997.

[27] Gibbs, G.R, Analysing qualitative data. Sage, London, UK, 2008.

[28] Graneheim U.H, and Lundman B. "Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness", Nurse Educ Today. 24(2):105-12. Feb, 2004.

[29] Guba, E.G, "Criteria for assessing the trustworthiness of naturalistic inquiries," Educ. Comm. Tech., 29(2):75–91, 1981.

[30] Hossain, E., Babar, M.A., Paik,H. "Using Scrum in Global Software Development: A Systematic Literature Review," Proc. of the 4th IEEE ICGSE, pp. 175-184, 2009.

[31] Khan, S.U., Niazi, M., and Ahmad, R. "Critical Success Factors for Offshore Software Development Outsourcing Vendors: A Systematic Literature Review," Proc. of the 4th IEEE ICGSE, pp. 207-216, 2009.

[32] Khurum, M. and Gorschek, T. "A systematic review of domain analysis solutions for product lines". J. Syst. Softw. 82(12):1982-2003, December, 2009.

[33] Kitchenham, B.A, and Charters, S., "Guidelines for performing Systematic Literature Reviews in Software Engineering", Version 2.3, Keele University, EBSE Tech. Report, EBSE-2007-01, 2007.

[34] Kitchenham, B.A, Dybå, T., and Jørgensen, M., "Evidence-based Software Engineering," Proc. ICSE'04, 273–281, May, 2004.

[35] Lofland, J., Analyzing Social Settings: A Guide to Qualitative Observation and Analysis. Belmont, CA: Wadsworth, 1971.

[36] Miles, M.B, and Huberman, A.M, Qualitative data analysis: An expanded sourcebook (2nd ed.). Thousand Oaks, CA: Sage, 1994.

[37] Noblit, G.W, and Hare, R.D, Meta-ethnography: Synthesizing Qualitative Studies, Sage, 1988.

[38] Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, N. Britten, M. Rodgers, K. Roen and S. Duffy. "Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Final Report". Swindon: ESRC Methods Programme, 2006.

[39] Saldaña, J.. The Coding Manual for Qualitative Researchers. London: Sage, 2008.

[40] Shull, F., Cruzes, D.S, Basili, V., Mendonca, M., "Simulating families of studies to build confidence in defect hypotheses", Inf. and Software Technology, 47(15):1019-1032, Dec. 2005.

[41] Staples, M. and Niazi, M. 2008. "Systematic review: Systematic review of organizational motivations for adopting CMM-based SPI," Inf. Softw. Technol. 50 (7-8): 605-620, June, 2008.

[42] Thomas, J., and Harden, A., "Methods for the thematic synthesis of qualitative research in systematic reviews", BMC Medical Research Methodology, 8:45, 2008.

[43] Walia, G.S. and Carver, J.C. "A systematic literature review to identify and classify software requirement errors". Inf. Softw. Technol. 51 (7):1087-1109, July, 2009.

[44] Williams, B.J. and Carver, J.C. "Characterizing software architecture changes: A systematic review". Inf. Softw. Technol. 52(1): 31-51, Jan. 2010.