Western University
Scholarship@Western

Electrical and Computer Engineering Publications   Electrical and Computer Engineering Department

9-2013

Using Meta-Ethnography to Synthesize Research: A Worked Example of the Relations between Personality on Software Team Processes

Fabio Q. B. Silva Dr.
Federal University of Pernambuco, fabio@cin.ufpe.br

Shirley S. J. O. Cruz
Federal University of Pernambuco, ssj@cin.ufpe.br

Tatiana B. Gouveia
Federal University of Pernambuco, tbg2@cin.ufpe.br

Luiz Fernando Capretz
University of Western Ontario, lcapretz@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/electricalpub

⊙ Part of the Computer Engineering Commons, Electrical and Computer Engineering Commons, Experimental Analysis of Behavior Commons, Personality and Social Contexts Commons, and the Software Engineering Commons

Using meta-ethnography to synthesize research: A worked example of the relations between personality and software team processes

Fabio Q. B. da Silva, Shirley S. J. O. Cruz, Tatiana B. Gouveia
Centre for Informatics
Federal University of Pernambuco
Recife, Brazil
{fabio,ssj,tbg2}@cin.ufpe.br

Luiz Fernando Capretz
Department of Electrical & Computer Engineering
University of Western Ontario
London, Canada
lcapretz@uwo.ca

Abstract— **Context:** The increase in the number of qualitative and mixed-methods research published in software engineering has created an opportunity for further knowledge generation through the synthesis of studies with similar aims. This is particularly true in the research on human aspects because the phenomena of interest are often better understood using qualitative research. However, the use of qualitative synthesis methods is not widespread and worked examples of their consistent application in software engineering are needed. **Objective:** To explore the use of meta-ethnography in the synthesis of empirical studies in software engineering through an example using studies about the relations between personality and software team processes. **Methods:** We applied the seven phases of meta-ethnography on a set of articles selected from a previously developed systematic review, to assess the appropriateness of meta-ethnography in this domain with respect to ease of use, and usefulness and reliability of results. **Results:** Common concepts were identified through reading and interpreting the studies. Then, second order translations were built and used to synthesize a model of the relationships between personality and software team processes. **Conclusions:** Meta-ethnography is adequate in the synthesis of empirical studies even in the context of mixed-methods studies. However, we believe that the method should not be used to synthesize studies that are too disparate to avoid the development of gross generalizations, which tend to be fruitless and are contrary to the central tenets of interpretive research.

Keywords— research synthesis; meta-ethnography; personality; team processes; software engineering.

I. INTRODUCTION

In this article, we were primarily interested in demonstrating the use of meta-ethnography [31] in the synthesis of empirical studies in software engineering. Broadly, research synthesis is a term used to describe a family of methods for summarizing, integrating, combining, and comparing the results of different studies which are mainly interested in similar or related research questions or topics [9]. Meta-ethnography [31] is an interpretive approach to research synthesis in which the primary concern is "to achieve synthesis through subsuming the concepts identified in the primary studies into a higher-order theoretical structure" [9]. This approach to synthesis is in contrast with aggregative approaches, like meta-analysis [14], that aim at summarizing data for the purpose of achieving generalizations [6].

Our motivation stems from the necessity of useful and reliable methods of interpretive synthesis when dealing with evidence from qualitative or mixed-methods empirical studies, as discussed by Cruzes and Dybå [9]. Meta-ethnography has its roots in social sciences [35], and has been applied successfully in health science [3] and other areas [27]. In software engineering, there are only two uses of meta-ethnography in systematic literature reviews, according to recent tertiary studies [9][10]. As far as we are aware, there is no published work that attempts to provide a worked example of the use of meta-ethnography to synthesize studies of different nature (qualitative and quantitative) in software engineering.

Therefore, our main goal is to show the process and the artefacts deployed in the development a meta-ethnographic synthesis of studies that used different research methods (quasi-experiments and ethnographies). We hope that our example will be helpful to researchers that are attempting to build interpretive synthesis of empirical studies either as part of systematic reviews or in theory building from multiple qualitative studies.

According to Noblit and Hare, the starting point of a meta-ethnography "involves identifying an intellectual interest that qualitative research might inform" [31]. We were, therefore, also interested in finding a theme that was worthy the effort of building the synthesis, that is, that could be of interest to us as researchers and also relevant for other researchers and practitioners. Because one of our main research interests is on the study of the social and individual aspects of software development teams, we decided to use the following research question to guide our meta-ethnography:

How does individual personality of team members relate with team processes in software development teams?

This article is structured as follows. In Section II, we briefly provide the conceptual background that is important to understand the studies and their synthesis. In Section III, we describe the phases of meta-ethnography, detailing the procedures and the artefacts used in each phase. In Section IV, we present the results of the meta-ethnography. In Section V, we discuss our results regarding the use of meta-ethnography and the results of the synthesis. Finally, we present concluding remarks in Section VI.

II. BACKGROUND

Meta-ethnography is described in in Section III. In this current section, we briefly present the conceptual background about the themes addressed in our synthesis. In this respect, some authors tend to make distinctions between teams and groups [18] while others use team and group to refer to the same concept [5]. A deeper discussion about this issue is out of the scope of this study and we will use the term software team or simply team, as its use is more widespread in software engineering literature.

A. Personality and Personality Tests

The research about personality is built around a diversity of theoretical foundations in the field of psychology, including traits, types, behavioural, and psychoanalytic theories. Of those traditions, traits and types theories are among the most used in organizational psychology [2] and in the studies about personality in software engineering [8]. The studies used in this synthesis focused on these two traditions.

Most of the studies about personality in software engineering use personality tests to identify differences among individuals [8]. In psychology there are two major categories of personality tests: projective and objective. Projective tests assess individual personality through responses from ambiguous stimulus with the assumption that personality is unconscious and the individual's responses will reveal its inner characteristics. Objective tests measure personality by self-assessment questionnaires with the underlying assumption that personality is primarily conscious and can be directly accessed. All studies synthesized in this article use some form of objective tests.

B. Individual Personality and Team Composition

While personality tests have been used successfully with individuals, its power can be multiplied when applied to teams [30]. Each person's personality traits can be used to help teams work more productively to accomplish common goals. Therefore, team composition has been studied in terms of diversity of the individual personalities within the team with at least the following three purposes: first, to help understanding how individual behaviour affect other within a team; second, to build highly productive team; and third to find ways to minimize conflict amongst team members. It is important to investigate how individuals prefer to respond to team challenges and the individual ways to interact with other on the team. Work on this field specifically aids team members by: identifying areas of strength and weakness for the team, clarifying team behaviour, helping to match specific tasks with individual skills, and providing a better framework to handle conflicts [20][30].

C. Team Processes

The term *team process* is broadly used to refer to what happens in the team or among teams, particularly in terms of the relationships between and amongst team members. According to Fuhriman et al. [12], team processes can be characterized and studied as phenomenon or as interaction. Seeing team process as phenomenon, or a composite of phenomena, is common in the research on teamwork. In this approach, processes, such as cohesion and conflict, are measure at certain points during the experimental period, averaged across subjects, and then presented statistically. Two studies in our synthesis used this approach [1][24].

A different approach is to view team processes as interactions, in which processes are analysed on both verbal and non-verbal content level. This approach leads to descriptively richer results with more explanatory power. However, measuring interactions is more time consuming and costly, and data analysis is more difficult. The preferred and most adequate form of data collection is through observations of team interactions, although some forms of questionnaires can also be used. This approach is not common in teamwork research, although it was the approach used to some extent in two studies in our synthesis [22][23].

The studies synthesized investigated cohesion and conflict as the main team processes of interest. Therefore, we briefly discuss these concepts below.

1) Cohesion

Cohesion can be defined as a dynamic process that reflects the willingness of group members to pull together and stay together as a unit to achieve their common goals and/or fulfil the affection needs of the individuals [4]. Carron and Brawley [4] put forward a model that identifies two dimensions to gauge cohesion (group integration and individual attraction to the group) that can be further divided into task and social orientation. According to their model, this measure of cohesiveness is neither evenly distributed across different teams nor with the same intensity along the teams existence. These authors assert that a team can be highly cohesive to accomplish a particular task despite conflicts among the members. Conversely, it is also possible that a socially cohesive team is unable to come together as a unit to a get a job done. The complex relationships between cohesion and conflict and team performance were clearly observed in our synthesis.

We shall refer to these two orientations as task and social cohesion. Although none of the synthesized studies used a multidimensional measure of cohesion, we found that the distinction between task and social cohesion provides a better explanation of the interpretations and translations in our synthesis.

2) Conflict

Intra-group conflict is broadly understood as perceived incompatibilities or the perceptions by the team members that they have discrepant views or have interpersonal incompatibilities [19]. Acuña et al. [1] conceptualized conflict as "an opposition or discrepancy between the ideas, beliefs or interests of the team members". Conflict can be classified in two types: task conflict and social conflict [19]. The task conflict occurs when members disagree about a decision or a task, presenting different ideas, opinions or points of view. Social conflict, in turn, happens when members exhibit incompatible values, tastes or ideas that generate wear personal between members [19].

The characterization of conflict in these two dimensions is important in the study of its relationships with other team processes, as cohesion, and how these relationships shape teamwork and determine some of its outcomes.

III. METHOD

We constructed our meta-ethnography using the seven phases suggested by Noblit and Hare [31]. In Section III.A, each phase is defined and the specific procedures we adopted are described in some detail. In Section III.B, we discuss threats to validity of our study and how we addressed them.

A. The Phases

1) Getting started

In this phase, the meta-ethnographer must look for a topic or problem that could be informed by an interpretive synthesis of a set of studies. The topic should be relevant to inform research and practice. Interpretive or qualitative approaches are the preferred research strategies when "how" and "why" questions are being asked [28]. Therefore, meta-ethnography aims at producing explanations, through the interpretation of findings from several studies, which were not evident in any of the individual studies. At this phase, a relevant topic of interest should be described by an adequate formulation of a research question.

As explained in the Introduction, our primary goal was to demonstrate the use of meta-ethnography as a method to synthesize empirical studies in software engineering. We also wanted to perform a synthesis that could be relevant and useful for researchers and practitioners interested in the study of software teams. Further, our own research interests are focused on understanding the role of individual and social factors in software development. Therefore, as a result of this phase, we proposed to answer the following research question: *How does individual personality of team members relate with team processes in software development teams?*

2) Deciding what is relevant to the initial interest

Noblit and Hare [31] emphasized that a meta-ethnography should not attempt to produce gross generalisations across studies with disparate goals and from too distinct contexts. In this sense, it goes in opposition to aggregative methods, such as meta-analysis or integrative reviews, which look for developing synthesis of an exhaustive list of studies, attempting to increase external validity of the results. In meta-ethnography, the translations and interpretations can be generalized, but as for all qualitative and interpretive studies, the context plays a central role in the generalization.

TABLE I. INITIAL SET OF SELECTED STUDIES

Id.	Research Method	Team Management	Ref.	Study Year[1]	Quality (0 – 11)
TP1	Quasi-experiment	XP teams	[1]	2004/5	9
TP2	Quasi-experiment	Self-regulating	[24]	N.I.	8
TP3	Ethnography	XP teams	[22]	2004/5	6
TP4	Ethnography	XP teams	[23]	2004	7
TP5	Survey	N.I.	[33]	N.I.	4

N.I.: Not Informed in the paper

We chose the studies of interest from the results of a previously developed systematic review [8] and its unpublished extension performed in 2011. In these reviews, we applied thematic analysis to classify 63 selected empirical studies into

---

[1] This date refers to the year in which the study was conducted, not when the article was published.

---

6 themes. To find the themes, we coded the variables, factors, and concepts from the research questions, hypothesis, and goals of each study. Five studies were classified into the theme *Team Process* because they addressed the relations between personality and one or more team processes. They formed our initial set of studies, shown in TABLE I.

We performed an initial screening on these studies to check if they formed a coherent set. Three of the studies were performed in 2004/5, and all of them used students as subjects. The style of team management was autonomous or self-regulating in all studies in which this information was available, with three of the studies investigating eXtreme Programming (XP) teams. The research methods used were different (two quasi-experiments, two qualitative studies with ethnographic characteristics, and one survey) but the objectives and central concepts addressed were fairly similar.

We then assessed the quality of the studies using the same criteria used by Dybå and Dingsøyr [11]. Two researchers assessed each study and the few disagreements were resolved in a consensus meeting. The two quasi-experiments received the highest scores, the ethnographies scored just above the average, and the survey scored lower the average. We decided to exclude TP5 because of its low score in the quality assessment. After removing TP5, considering the similarities of the four remaining studies, we concluded that the studies formed a coherent set adequate for a meta-ethnography.

3) Reading the studies

This phase involved carefully reading the papers to gain a general view of the set of studies. Further, the goal was to identify the key concepts addressed in each individual study through repeated reading and noting of the main concepts. However, the name of this phase is perhaps misleading because we actually read the studies many times in the following phases as well. In each phase, the emphasis of the reading changed, as the researchers became more familiar with the studies and deeper understanding and interpretations were needed. Noblit and Hare admitted that, because in qualitative research the synthesis develops through the synthesis effort, "this phase is not so clear" [31].

In this phase, three important tasks must be performed. First, the careful reading of all studies to create an initial understanding of the whole. Second, the extraction of contextual information from each study that would be important in the interpretations and translations in future phases. For this, we contend that the type of contextual information must be defined *a priori*, depending on the information needed in the specific synthesis. Further, this extraction should be carried out by at least two researchers to increase reliability and disagreements should be noted and properly addressed. Third, relevant concepts associated with the research questions should be identified in each study.

In our study, we started by reading the four studies without considering data extraction. Then, we extracted contextual data in pairs. Disagreements in the extractions were resolved in consensus meetings. The result of this extraction is presented in TABLE II. Next, we read the studies for a third time looking for the main concepts that were related to our research question and this resulted in TABLE III.

4) Determining how the studies are related

In this phase, the relationships between the different studies were considered. It was the first phase in which we put the studies together and started making comparisons among them. Noblit and Hare [31] suggested juxtaposing the lists of key metaphors, ideas, and concepts occurring in each study as the initial step in relating the studies.

We identified three questions that guided our work in this phase:

• *Are the studies addressing the same or similar concepts?*

• *Are the operational definitions of the concepts similar across studies?*

• *Are the studies addressing similar relationships among concepts?*

To answer the first question, we used TABLE III and identified the concepts that were addressed by at least two studies. Six concepts from TABLE III fulfilled this criterion. We considered that it would only make sense to synthesize results related to concepts that were addressed by at least two studies; otherwise no synthesis would be made. Then, to address the second question, we looked for the conceptual and operational definitions of the six concepts: personality, cohesion, conflict, team composition, satisfaction, and team performance. These definitions can be found in TABLE IV.

We performed another reading of the studies looking for the relationships among concepts that were addressed in each study. We identified the relationships and extracted the results of the studies in TABLE V. Tables II, IV, and V express the similarities and differences among the studies, and are the result of this phase.

5) Translating the studies into one another

During this phase, we translated the concepts and relations from one study into the concepts and relations of the other studies. In meta-ethnography, we start by considering studies as analogies, that is, considering that findings in one study are like findings in the other studies, but also taking into account their non-similarities. However, translations go beyond simple analogies in that specific meanings from each study are preserved and compared across studies through the synthesis process. An adequate translation preserves the meanings of concepts and their relations in each study. It also compares the meanings of concepts and relations from on study with concepts and relations from the other accounts. In general, concepts and relations from the studies can compare with each other in three ways: they are directly comparable as *reciprocal* translations; they may contradict or stand in opposition to one another as *refutational* translations; or taken together they may represent a *line-of-argument* [3][31].

We used Table IV as the input to the translations. We also recurred to the studies whenever we needed to get a deeper understanding of the concepts and relations. We translated each row of Table IV into a first-order synthesis of the relations and used these translations as the input to phase 6.

6) Synthesizing translations

Noblit and Hare [31] discussed that, when the number of translations is large, it is desirable to synthesize them into another second-order of synthesis, "making a whole into something more than the parts alone imply" [31]. We produced the synthesis of the translations and put them together with the first-order translations in TABLE VI. We then created a diagram that summarizes this synthesis and a narrative that describes the its central story, in a similar way as a central story is developed in grounded theory (Fig. 1).

7) Expressing the synthesis

This article is one attempt to report the synthesis result. We used a structure similar to Britten et al. [3]. The primary audience is the research community interested in performing synthesis of empirical research using meta-ethnography. We also believe that our synthesis can inform researchers and practitioners interested in understanding how personality relates to team processes in software teams.

B. Threats to Validity

Because meta-ethnography is an interpretive approach to synthesis, we addressed validity and reliability of our synthesis from the three perspectives proposed by Merriam [28].

Credibility or Internal Validity: in interpretive accounts, internal validity is related to the issue of whether the research results consistently represent reality. In a synthesis, internal validity relies on the credibility and internal validity of the synthesized studies. We addressed this issue by carefully scrutinizing the quality of the primary studies in phase 2. We only chose studies that score above the average and also scored well on issues related to study design and development. This is an important because synthesizing low quality studies will inevitably result in a low quality synthesis no matter how carefully and consistently the synthesis was performed. This issue must be addressed mainly in phases 1 and 2, but also in phase 3 new problems can be found that affect internal validity.

Consistency or Reliability: An important question in qualitative research is whether the findings are consistent with the data collected. In a meta-ethnography, the question is whether the translations and the interpretations are consistent with the concepts, metaphors, and results in each study. Preservation of the meanings from each individual account is at the core of meta-ethnography [31]. As we progress from phase 3 to phase 6, our attention moved from the individual accounts to the translations between them. Therefore, the consistency of the data extracted in phase 3 and the relations between studies in phase 4 are very important because the translations in phase 5 are based on them. We performed phases 3 and 4 in pairs, discussing disagreements and reaching consensus after deep debate among researchers. The first author performed phases 5 and 6 and the other authors reviewed the results looking for potential inconsistencies in the higher-level interpretations and in the final synthesis.

Transferability or External Validity: It is a common understanding in qualitative research that it is the reader or user of the study that should primarily engage in the generalization of research findings [28]. In this sense, the reader or user can decide to what extent the findings can be applied to other situations. The researcher has to enhance the possibility of someone else "transferring" the results and this is accomplished by a rich description of the context in which the

research was carried out. In a meta-ethnographic synthesis, the contexts are those from the synthesized studies. In our synthesis, we worked to enhance transferability in two ways. First, we chose studies from similar or related contexts. Second, we extract and presented contextual information in TABLES I, II and IV so the readers can quickly assess and compare the contexts of the studies with their own context.

IV. RESULTS

This section is structured following the phases described in Section III. The results of phases 1 and 2 were presented before and, thus, in this section we describe the results of phases 3-6. In the tables that summarize the results, sentences between double quotes are literal transcriptions from each study.

A. Results from phase 3: Reading the studies

We collected data about study objective and aspects of the study design and development to enable comparisons and also to make sense of the translations and interpretations (TABLE II). Three studies used a test related to Myers-Briggs Type Indicator (MBTI) to access individual personality and only TP1 used a version of NEO-PI based on the Five Factor Model. TP1 and TP2 were quasi-experiments and TP3 and TP4 were meta-ethnographically informed qualitative researches. Consistently, data used in the quasi-experiments were quantitative and collected through the application of questionnaires in certain

points during the experimentation process, whereas observation was used as the main data collection technique in the qualitative studies. In this sense, TP1 and TP2 used a process as phenomena approach, whereas the other studies employed an approach closer to process as interaction, as discussed in Section II.

Another important similarity among the studies is that all of them investigated teams of students in university level courses using some type of autonomous team (XP teams in three studies). Although this can apparently restrict generalizations to other contexts, from an interpretive stance this in fact produces a deeper understanding of the phenomena in this specific context. From this contextual information, we concluded the studies were sufficiently diverse in content and type of data to produce rich interpretations and yet were not so disparate to allow a consistent synthesis.

We then read each study again looking for the main concepts related to our research question. As the studies investigated several different aspects of teamwork, in particular, TP3 and TP4, it was important to use the research question to keep the focus of our readings. We extracted eight concepts related to teamwork and Table III shows in which study they were addressed. The contents of TABLES II and III summarize the results of phase 3.

TABLE II. CONTEXTUAL INFORMATION ABOUT THE STUDIES

Context	TP1 [1]	TP2 [24]	TP3 [22]	TP4 [23]
Objective	"This article analyses the relationships between personality, team processes, task characteristics, product quality and satisfaction"	"We test the impact of problem solving preferences (a sub-set of the MBTI scale) on group conflict and performance".	"… investigate interactions of personalities in software engineering (SE) teams and how disruptions and lack of debate between individuals affected performance".	"… to gain a qualitative understanding of how cohesiveness relates to personality type, performance and adherence to a methodology (XP)."
Sample	Second-year computing undergraduate students (105 participants divided in 35 teams)	Undergraduate students, enrolled in two 15-week SE courses. (38 members in 9 teams)	Three teams (5-6 individuals each) of Master's-level students.	Five teams (5-6 individuals each) of Master's-level students.
Research Method	Quasi-experiment	Quasi-experiment	Ethnographically-informed	Ethnographically-informed
Design	"The students were divided into 35 three-member teams … formed at random and … blind to the quasi-experimental conditions and hypotheses."	"… students were assigned to 4-5 person teams: five control groups of numerical dominant problem solving style and four experimental groups of diverse styles."	Convenience sampling of the three teams participating in the "Maxi Project".	"The teams were selected on the basis of personality type, nationality and previous skills/experience".
Data Collection	"Measurements were taken before the project (NEO FFI personality test), during the project (conflict, cohesion) and after the project (autonomy, interdependency and satisfaction)."	"At the conclusion of every phase of the team project, peer evaluations were collected. Team members were asked five questions related to team dynamics".	Observations and online personality test based on the MBTI.	Observations, focus group interviews, document analysis, workgroup cohesion test, and online personality test based on the MBTI.
Setting	"Special-purpose project with non-professional participants (… students) undertaking a (toy) project using an adaptation of the agile XP method within a laboratory environment".	"The semester long projects were complex and ill-structured, requiring teams to consider the pros and cons of several design options".	"The teams … worked on real software development projects for real clients in the project "Maxi Project" (a two semester long project during 2004-2005)".	Teams of students participating in professional software house known as Genesys Solutions as part of the Software Engineering Observatory at the University of Sheffield.
Country	Spain	United States	England	England

TABLE III.   MAIN CONCEPTS FROM EACH STUDY

Concepts	TP1	TP2	TP3	TP4
Task Characteristics	X			
Personality	X	X	X	X
Conflict	X	X		X
Cohesion	X			X
Team Composition		X	X	X
Performance		X	X	X
Satisfaction	X	X		
Software Quality	X			

B. Results from phase 4: Determining how the studies are related

We looked for similarities and discrepancies among the studies to guide the construction of the translations in phase 5. As discussed above, task characteristics and software quality were not analyzed because they were addressed in a single study. Our next concern was to identify the theoretical and operational definition used in each study for the remaining six concepts (TABLE IV).

Personality was clearly defined both at the theoretical and operational levels in all four studies. Three studies used objective tests based on MBTI and TP1 used a version of NEO-FI test. MBTI [29] is based on the typological theory of personality developed by Carl G. Jung [21], but none of the studies in our synthesis used the official version of MBTI. NEO-FI test is based on the Five Factor Model of personality traits [7], and TP1 used the official Spanish version of the test.

Cohesion scales were used in two studies. TP1 used the Gross Cohesion Scale [15], which is self-report measure with 9 items considered to be one-dimensional. TP4 used the workgroup cohesion scale developed by Price and Mueller [34], which has 8 items and is also one-dimensional. TP1 applied the cohesion questionnaire in the middle of the project whereas in TP2 the questionnaire was applied at the beginning, middle, and end of the project and the average was used to determine the workgroup cohesion.

Operational definitions of conflict were used in TP1 and TP2 and both used Jehn's definition of intra-group conflict [20], which is a two-dimensional scale that measures social and task conflicts. In TP4, task conflict was observed and analysed during the study, but no operationalization was used to achieve a quantitative measure of conflict.

Team composition was not directly defined in TP1, but the aggregation of individual personality traits to a "team personality" is close to notions of composition based on personality types that were used in the other studies. TP2 used team composition explicitly in the quasi-experimental design, considering homogeneous and heterogeneous teams with respect to the problem solving preferences, a sub-scale of MBTI. The remaining studies also considered composition diversity in terms of the personality types in the team, but did not address any particular type of composition.

Satisfaction was studied in TP1 and TP2 as a measure of outcome in the teamwork. TP1 used a three-item scale from Gladstein [13] and TP2 used a six-item scale from Pershall and Ellis [32]. Although both operationalizations are different, the scale from Pershall and Ellis [32] has its theoretical basis on the work of Gladstein [13], suggesting similarities at the theoretical level. Finally, all studies considered the project or course grade as a measure of team performance. However, the operationalization of this measure was not clearly presented in any study.

Considering the differences of the operational definitions of personality, cohesion, conflict, and satisfaction, it would not be feasible to integrate the results at the operational level if we were using an aggregative synthesis approach. Using an interpretive approach we could compare the differences among studies at the conceptual or theoretical level and still arrive at consistent interpretations.

After comparing the concepts across the studies, we identified and compared the relationships between concepts (TABLE V). Team composition, defined in terms of the personality of team members, was the central antecedent factor addressed in all four studies. Relationships between composition and conflict were found in TP2, TP3, and TP4, which also found direct relationships between composition and team performance. TP1 and TP4 found relationships between composition and cohesion. In particular, TP1 found that teams with high levels of Extraversion and Agreeableness presented high levels of cohesion. TP1 and TP4 also found relationships between the team processes cohesion and conflict. In these studies, high cohesion in certain teams tend to reduce conflict, whereas other teams with high levels of social conflict showed low levels of cohesion.

No direct relationship between cohesion and outcomes such as performance and satisfaction was found in TP1, but TP4 found that cohesive teams tend to outperform teams with low cohesion. This suggests that other factors act as intermediates between cohesion and team outcomes, and we proposed to use Effort Applied to the Task as one such a factor as in Hackman's theory [16]. Finally, social conflict was clearly related to low levels of performance and satisfaction in TP1 and TP2, and TP4 identified that certain levels of task conflict were favourable in forcing the teams to evaluate different alternatives to approach problems during the development of the projects. These results about conflict suggested that social and task conflict played distinct and potentially opposing effects in the results of teamwork.

C. Results from phase 5: Translating the studies into one another

We started the translation between studies when we identified the relationships between the main concepts and built TABLE V in phase 4. In phase 5, we built interpretations of all cells in a given row of TABLE V, and created first-order translations of them. In this process, we produced lines-of-argument consistent with the individual accounts, preserving the meanings of concepts from each study. These translations are presented in TABLE VI.

D. Results from Phase 6: Synthesizing the translations

In phase 6, we synthesized the first-order translations produced in the previous phase, creating second-order translations with the goal of making a whole and coherent account of the synthesized studies.

TABLE IV.   CONCEPTS DEFINITIONS AND OPERATIONALIZATIONS

Concepts	TP1 [1]	TP2 [24]	TP3 [22]	TP4 [23]
Personality Theory	Five Personality Factor or Big Five	Myers-Briggs Type Indicator (MBTI)	Myers-Briggs Type Indicator (MBTI)	Myers-Briggs Type Indicator (MBTI)
Personality Test	Spanish version of the NEO FFI	Assessment to identify problem-solving style (inner-dimensions of the MBTI).	An online test based on MBTI developed by Human Metrics [17].	An online test based on MBTI developed by Human Metrics [17].
Cohesion - Definition	''A group property with individual manifestations of feelings of belongingness or attraction to the group''.	_	_	(The paper addresses cohesion, but does not provide a definition of the concept)
Cohesion - Operationalization	Gross Cohesion Questionnaire [36].	_	_	Price and Mueller scale [34].
Conflict	"Conflict is an opposition or discrepancy between the ideas, beliefs or interests of the team members. … [19]"	Intragroup conflict, which encompasses task and relationship conflict as defined by Jehn et al. [20].	_	Task-related conflict refers to disagreement among the team members about task issues, including the nature and importance of task goal.
Conflict - Operationalization	"Intragroup Conflict questionnaire from Jehn [19] measures both task conflict and social conflict."	Questionnaire with questions drawn from the intragroup conflict scale developed by Jehn [19].	_	(Conflict is not operationalized)
Team Composition	(It is not directly addressed, but the aggregation of individual scores to a "team personality" is close to a notion of composition based on personality types).	Composition diversity defined in terms of the dominance of problem-solving style.	Diversity of personality types and ethnicity in the teams.	Diversity of personality types in the team
Satisfaction	Gladstein [13].	Pershal and Ellis [32]	_	_
Performance	Course grade	Team's grade	Not defined	Project Grade

TABLE V.   RELATIONSHIPS BETWEEN CONCEPTS

Relations	TP1 [1]	TP2 [24]	TP3 [22]	TP4 [23]
Composition and Conflict	_	"Problem solving dominant groups showed higher levels of intra-group conflict".	"This team work … well … despite great differences in both personality and ethnicity".	"… agreement is straightforward if one has a team of almost homogenous personality types".
Composition and Cohesion	Team Extraversion and Agreeableness relate positively with Cohesion.	_	_	The results indicate that certain teams … were found to be very cohesive due to a mixture of [personality] types.
Composition and performance	_	Diversity in problem solving style correlates positively with performance	"… it takes a variety of skills and personalities to solve the myriad of problems related to SE".	"This range of personalities was a source of strength … because the different types came up with different ways of approaching the same problem".
Cohesion and Conflict	"Level of cohesion drops the greater task and social conflict are among team members."	_	_	"… there was almost too much emphasis on not offending people. This meant that work was sometimes not done and there were no recriminations".
Cohesion and Performance	Cohesion is not related to Satisfaction and Performance	_	_	"Teams with a high level of cohesion tend to outperform other team with lower levels".
Conflict and Performance		"Significant negative relationship between the Intra-Group Conflict and Grade".	_	"The task conflict in this case seems to have more favourable effect on the performance of the team".
Conflict and Satisfaction	"Level of satisfaction drops the greater task conflict is among team members."	"Homogeneous groups showed higher levels of intra-group conflict and lower satisfaction".	_	_

We then created TABLE VI with three columns. The first summarized our interpretation of the main findings of each study that were directly related to our synthesis. The second and third rows present the first and second order translations of the studies. We used the translations to build a central narrative that represents our synthesis, illustrated in Fig. 1.

1) The Central Narrative of the Synthesis

Individual personalities in teams have an influence on its composition. Certain combinations of personality types will tend to favour certain types of interactions among team members, as also discussed by Nash [30]. Therefore, levels of cohesion and conflict will be related to the type of personality composition in the team. We found evidence of distinct influences of two (proposed) archetypical team compositions: one that favours social interactions and cohesion in the team (we called this composition as Pro-SCH, for pro-social cohesion), and another that is focused on getting the job done (we called it Pro-TSK, for pro-task).

The relationships between conflict and cohesion are complex, with mutual cycles of reinforcement. A team exhibiting high levels of social cohesion (Social-CH) will favour interactions that tend to decrease conflict. In such a situation, the social interactions will flourish and satisfaction tends to increase. On the other hand, a team exhibiting high levels of social conflict (Social-CF) will tend to enter in a pattern of behaviours that will decrease cohesion. From the accounts of the studies, it seems that which reinforcement cycle will prevail is mediated by the type of team composition. For instance, from TP1, it seems that a team with high concentration of Extraversion and Agreeableness will tend to favour the increase-cohesion-decrease-conflict cycle. From TP2, teams with dominant problem-solving styles will enter more easily on the increase-conflict-decrease-cohesion cycle.

Cohesion and conflict seems to have a moderating effect on the effort individual team members apply to the team tasks. This is consistent with Hackman's theory of team effectiveness [16]. In our synthesis, we found that cohesion seems to positively affect task effort while social conflict will have the opposite effect. Task effort was not directly addressed in any study and we added this factor in our synthesis as a proposition that requires verifications.

However, these effects are not straightforward because highly cohesive teams can also tend to have sub-optimal performance because of teams members will avoid confrontations that sometimes are necessary to avoid loafing and to reach better solutions through an honest and open assessment of alternatives (this effect is known as the Abilene paradox [25]). In such situations, teams can experience high satisfaction with their work, due to absence of conflict, and still perform poorly. Further, as found in TP4, certain levels of (task) conflict are important for performance because better solutions were found. We modelled this effect by showing that (social) cohesion is likely to reduce the strength positive relationship between task conflict and task effort.

V. DISCUSSION

A. About the use of meta-ethnography

Our main conclusion and warning is that meta-ethnography is not straightforward to use. The several readings of the studies, data extraction performed in pairs, reaching agreements, and checking potential inconsistencies in the interpretations and translations are time consuming and require maturity in the research team. Experience with the methodology itself and a deep understanding of the philosophical stances that form the cornerstones of interpretive research are required. Because the method requires the careful in preservation of the meanings of concepts and metaphors from the synthesized studies, reliability is dependent on the commitment of researchers to respect the findings of the individual accounts and not making generalizations that are not supported by them.

Scalability is also an important issue. We think that synthesizing too many studies is not practical due mainly to the amount of translations and interpretations. Other researchers (in personal communications) have proposed to perform incremental synthesis starting with two studies and then proceeding by adding the remaining studies to the previous synthesis. We belief this approach should be used with extra-caution because it would be very easy to forget the meanings of the previously synthesized studies as the synthesis proceeds, with the danger of subverting their meanings and resulting in syntheses with low reliability.

B. About the result of our synthesis

We believe that our model improved the understanding of the studied phenomena and provides a contribution mainly to researchers but also to practitioners interested in understanding the complex relationships in software teams. The resulting model of Fig. 1 and the corresponding narrative are propositions built from the evidence but that still require further verification. In particular, as part of our synthesis we proposed that certain relationships between factors are causal. However, all four studies only provided evidence of correlations between them. The causal relationships proposed still need empirical tests.

Further, our results are not transferable to any context. Our meta-ethnography is an account of the culture of software teams formed by students in the context of university courses. Maturity, and other factors not addressed in this context, will certainly influence the relations we described in our synthesis. It is, therefore, essential that studies in industrial settings with individuals with varying degrees of personal and professional maturity be performed to produce new evidences in contexts or to verify the applicability of the current propositions.

Finally, only TP1 explicitly used a theory to guide the definition of variables and hypotheses [26]. This made relating the studies and building the translations more difficult and more prone to inconsistencies. Further, the results have little explanatory power and consequently low reduced applicability. We strongly believe that a systematic research framework to investigate team process would be important to improve individual studies and support more useful and reliable synthesis, as suggested by Fuhriman [12].

TABLE VI.   SYNTHESIS: FIRST AND SECOND ORDER TRANSLATIONS

| Relations	Hypothesis or Conclusions	First-order Translations	Second-order Translations
Composition and Conflict	TP4 "combinations of personality types are important and they can have an effect both on performance and cohesion. ... a team whose members do not have very close friendly relationships with each other, can still be [task] cohesive and outperform others that enjoy more open relaxed relationships [social cohesion].	The interplay of composition diversity, conflict, and cohesion is complex. On the one hand, agreement is easier to achieve and homogeneous groups. On the other hand, certain types of homogeneous groups tend to have higher levels of intra-group conflict. This suggests type of dominance in the group is a moderating factor on these relationships.	Certain compositions tend to favour group social cohesion (Pro-SCH) while others tend to have a greater focus on performing their tasks (Pro-TSK). Pro-SCH teams will avoid conflict (task and social), which may increase satisfaction but reduce task efforts through its effect on task conflict. Pro-TSK teams will be more prone to have intra-group social conflict and reduced social cohesion due to their focus on get the job done at all costs, but this may increase task effort directly and also through levels of task conflict.
Composition and Cohesion	Such a team achieves this by virtue of concentrating on the task in hand and devoting all energies to completing a given project." TP2 "Teams with diverse problem-solving styles performed better … and also displayed more creativity, which can be explained by low levels of social conflict in such teams". TP3 ""... it takes a variety of skills and personalities to solve the myriad of problems related to SE".	A mixture of personality types that include extraversion and agreeableness seem to produce teams with higher cohesion and lower intra-group social conflict. Certain homogeneous compositions (related to inner-dimensions of MBTI) create teams that are more likely to have high levels of social conflict.	
Composition and performance		In task with high levels of problems to be solved diversity of personality types will produce more effective teams. Routine or non-problem-solving tasks seem to be better performed by homogeneous teams with traditional science/engineering personality.	Diversity in composition seems to indirectly be related with performance, through its relationship with cohesion and conflict, and task type has a moderating effect on these relationships.
Cohesion and Conflict	TP1- Cohesion is negatively correlated to task conflict and (social and task) conflict correlates negatively with Satisfaction. However, cohesion is not correlated with satisfaction and both cohesion and conflict are not correlated with performance, defined as software quality.	Cohesion and conflict relate in a complex feedback loop. On one hand, a very cohesive team will under perform due to the need to avoid social or inter-personal conflict. On the other hand, high levels of (task or social) conflict will reduce the bonds in the team, reducing cohesion.	The type of team composition seems to influence the balance between cohesion and conflict, with Pro-SCH teams tending to shift the balance towards social cohesion and Pro-TSK teams tending to have more intra-group social conflict.
Cohesion and Performance		Cohesion does not seem to be directly related to effectiveness outcomes, but increases the likelihood of team effectiveness through other moderating effects on the effort the team puts on the tasks.	Cohesion and conflict seem to have an indirect effect on performance through their effect on the effort the individuals in the team are willing to apply on the tasks. Cohesion and task conflict seem to have a positive effect on effort, whereas social conflict has a negative effect.
Conflict and Performance		Task conflict is favourable in the analysis of multiple alternatives in performing a team task whereas social conflict relates negatively with performance.	
Conflict and Satisfaction		Intra-group social-conflict reduces individual satisfaction with the team.	Satisfaction with the team and teamwork does not seem to be directly related to performance.
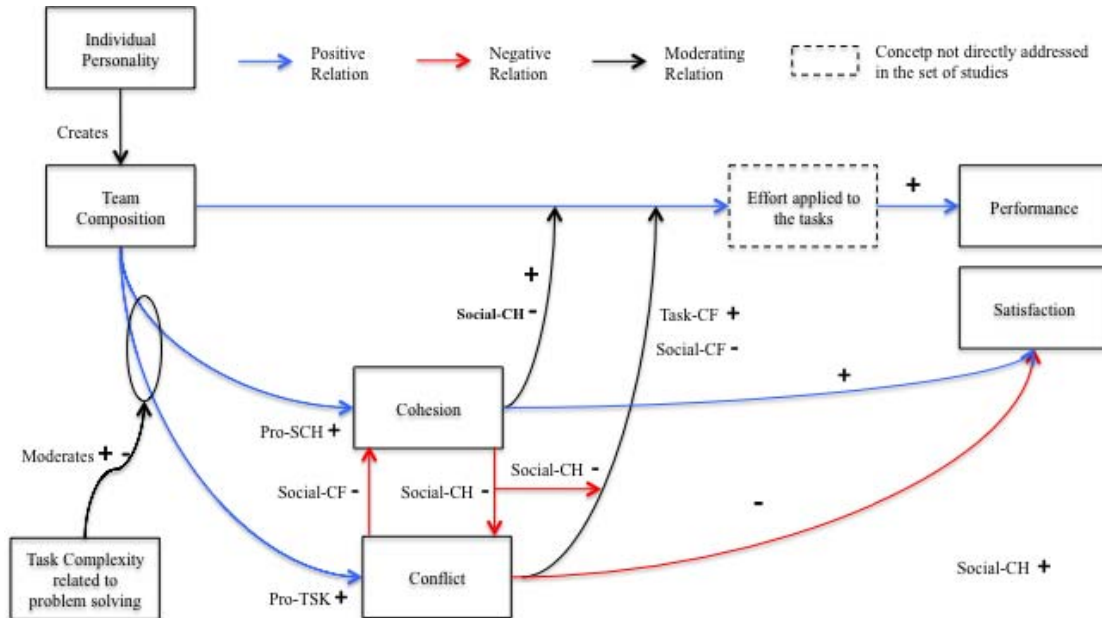
Fig. 1.   Synthesis of Translations: A model of the relationships between team composition and team factors in software teams

VI. CONCLUSION

We presented a worked example of a meta-ethnography to synthesize four studies that addressed the relationships between team member personality, team processes and outcomes of teamwork. We hope to have demonstrated that meta-ethnography is an adequate method to produce synthesis of mixed-method empirical studies and that the results produced in the synthesis are more than just the sum of the individual studies, but still preserve their meanings. We also hope to have achieved a synthesis of the studies useful to guide future studies in the theme.

Researcher engaging in the use of meta-ethnography must be aware that the set of studies to synthesize greatly influences the consistency and reliability of the resulting synthesis. The synthesis of bad studies will inevitably lead to bad results. Further, studies that are too disparate in objectives and context would tend to produce gross generalizations with limited consistency and usefulness.

Therefore, our experience seem to indicate that meta-ethnography is better suited for the synthesis of a small set of consistently related studies with respect to objectives, conceptual background, and contexts. In this sense, its application in systematic reviews with large number and broad range of studies should only be performed in subsets of studies in similar themes, as we did in our example.

ACKNOWLEDGMENT

REFERENCES

[1] S. T. Acuña, M. Gómez, N. Juristo, How do personality, team processes and task characteristics relate to job satisfaction and software quality?, Information and Software Technology Vol. 51, No. 3, 627–639, 2009. http://dxdoi.org/10.1016/ j.infsof.2008.08.006.

[2] N. Anderson et al. (Eds.). Handbook of industrial, work and organizational psychology, Volume 1: Personnel psychology. Thousand Oaks, CA: Sage 2002.

[3] N. Britten et al. Using meta ethnography to synthesise qualitative research: a worked example, Journal of Health Services Research & Policy Vol. 7 No. 4, 209–215, 2002.

[4] A. V. Carron, L. R. Brawley. Cohesion: Conceptual and Measurement Issues, Small Group Research, 31, 1, 89-106, 2000.

[5] S. G. Cohen, D. E. Bailey, What makes team work: group effectiveness research from the shop floor to the executive suite, Journal of Management 23, 239–290, 1997.

[6] H. Cooper, et al. (Eds.), The Handbook of Research Synthesis and Meta-Analysis, second ed., Russell Sage Foundation, 2009.

[7] P.T. Costa Jr., R.R. McCrae, NEO Personality Inventory, Revised, Psychological Assessment Resources, Odessa, Florida, 1992

[8] S. J. O. Cruz, et al. Personality in software engineering: preliminary findings from a systematic review, in: Proceedings of the 15th Annual Conference on Evaluation and Assessment in Software Engineering (EASE 2011), Durham, 2011. doi:10.1049/ ic.2011.0001.

[9] D. S. Cruzes, T. Dybå. Research synthesis in software engineering: A tertiary study. *Inf. Softw. Technol.* 53, 5, 440-455, 2011. doi:10.1016/j.infsof.2011.01.004.

[10] F. Q. B. da Silva, et al. Six years of systematic literature reviews in software engineering: An updated tertiary study. *Inf. Softw. Technol.* 53, 9, 899-913. 2011 doi:10.1016/j.infsof.2011.04.004.

[11] T. Dybå, T. Dingsøyr. Empirical studies of agile software development: a systematic review, IST 50 (9-10), 833–859. 2008.

[12] A. Fuhriman, S. Drescher, G. Burlingame. Conceptualizing Small Group Process. Small Group Research, 15: 427, 1984 doi: 10.1177/104649648401500401

[13] D. L. Gladstein, Groups in context: a model of task group effectiveness, Adm. Sci. Q 29, 499–517, 1984.

[14] G. Glass, Integrating findings: the meta-analysis of research, Review of Research in Education 5 (1977) 351–379.

[15] N. Gross & W. E. Martin. On group cohesiveness. *American Journal of Sociology,* 57: 546–564, 1952.

[16] J. R. Hackman, The design of work teams, in: J. Lorsch (Ed.), Handbook of organizational behavior, Prentice-Hall, Englewood Cliffs, pp. 315-342, 1987,

[17] HumanMetrics. http://www.humanmetrics.com/cgi-win/JTypes2.asp.

[18] J. Ivancevich, M. Matteson, Organisational Behaviour and Management, McGraw Hill, Irwin, New York, 2002.

[19] K.A. Jehn. A multimethod examination of the benefits and detriments of intragroup conflict, Adm. Sci. Q 40, 256–282, 1995.

[20] K. A. Jehn et al. "Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups", Administrative Science Quarterly, Vol. 44, 1999, pp. 741–763.

[21] C. G. Jung, Psychological Types *The Collected Works of C.G Jung,* Harcourt Press, 1923.

[22] J. S. Karn, T. Cowling. A follow up study of the effect of personality on the performance of software engineering teams. In: Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering. ACM, 232-241, 2006.

[23] J. S. Karn et al. A study into the effects of personality type and methodology on cohesion in software engineering teams. Behaviour & Information Technology, Vol. 26, No. 2, 99 – 111, 2007.

[24] T. L. Lewis. W. J. Smith. Creating high performing software engineering teams: the impact of problem solving style dominance on group conflict and performance. *J. Comput. Small Coll.* 24, 2, 121-129, 2008.

[25] J. McAvoy, T. Butler. "The impact of the Abilene Paradox on double-loop learning in an agile team". Information and Software Technology 49 (6): 552–563, 2007. doi:10.1016/j.infsof.2007.02.012.

[26] J. E. McGrath, Groups: Interaction and Performance, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

[27] C. H. Major, M. Savin-Baden. Integration of qualitative evidence: towards construction of academic knowledge in social science and professional fields. Qualitative Research 11: 645-663, 2011, doi:10.1177/1468794111413367.

[28] S. B. Merriam, Qualitative Research: A Guide to Design and Implementation. Jossey-Bass, San Francisco, 2009.

[29] I. B. Myers et al. MBTI Manual: A Guide to the Development and Use of the Myers–Briggs Type Indicator, 3rd Edition. Consulting Psychologists Press, Palo Alto, CA, 1998.

[30] S. Nash, Turning Team Performance Inside Out: Team Types and Temperament for High-Impact Results, Nicholas Brealey, 2010.

[31] G. W. Noblit, R. D. Hare, Meta-Ethnography: Synthesizing Qualitative Studies (Qualitative Research Methods). [S.l.]: Sage Publications Inc., 1988.

[32] M. J. Pearsall, A. P. J. Ellis. The Effects of Critical Team Member Assertiveness on Team Performance and Satisfaction. Journal of Management, Vol. 32 No. 4, pp. 575-594, 2006.

[33] A. R. Peslak. The impact of personality on information technology team projects. In Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research. ACM, New York, 273-279. 2006. DOI=10.1145/1125170.1125233

[34] J. L. Price, C. W. Mueller. Handbook for Organizational Measurement, Marsfield, USA, Scranton: Harper-Collins, 1986.

[35] M. Savin-Baden, C. H. Major. Using interpretive meta-ethnography to explore the relationship between innovative approaches to learning and innovative methods of pedagogica research. Higher Education 54(6): 833–852. 2007.

[36] J. P. Stokes. Toward an understanding of cohesion in personal change groups, Int. J. Group. Psychother. 33, 449–467, 1983.