This is a post-peer-review, pre-copyedit version of this article. The final authenticated version is available online at:
http://dx.doi.org/10.1145/3463274.3463841

Challenges when Applying Repertory Grid Technique for Software Practices

Diana Kirk
The University of Auckland
Auckland, New Zealand
diana.kirk@auckland.ac.nz

Kelly Blincoe
The University of Auckland
Auckland, New Zealand
k.blincoe@auckland.ac.nz

ABSTRACT

The *Repertory Grid Technique (RGT)* has been applied within the software engineering domain to investigate a variety of topics. These include topics relating to architectural knowledge, team level tacit knowledge, and project success mechanisms. The technique is based on *Personal Construct Theory (PCT)* and is claimed to be suitable for gaining a deep understanding of peoples' perspectives on a topic. The essence of RGT is a consideration of similarities and differences, for example, between different project instances. In this paper, we describe a case study in which we applied the technique with the aim of eliciting practitioners' viewpoints on contextual factors for situated software practices. We interviewed twelve practitioners in three organisations. We found that the RGT approach was challenging to implement for several reasons. Participants had difficulty in choosing specific instances of a software practice, identifying similarities and differences tended to be problematic and causal pathways were not always easy to establish. Our contributions are the highlighting of the challenges that may occur when implementing this technique, an analysis of the issues encountered and some possible mitigation approaches. These may serve as support for SE researchers considering an RGT based study.

CCS CONCEPTS

• **Software and its engineering → Software development methods**; *Programming teams.*

KEYWORDS

Personal Construct Theory, Repertory Grid Technique, Industry Study, Software Development Context,

ACM Reference Format:
Diana Kirk and Kelly Blincoe. 2021. Challenges when Applying Repertory Grid Technique for Software Practices. In *Evaluation and Assessment in Software Engineering (EASE 2021), June 21–23, 2021, Trondheim, Norway.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3463274.3463841

1 INTRODUCTION

Personal Construct Theory (PCT) was created in the 1950s by George Kelly based on the observation that events faced by individuals are "subject to as great a variety of constructions as our wits will enable us to contrive" [16]. Kelly named this philosophical perspective on 'reality' *constructive alternativism*. It reflects the notion that a person understands reality by interpreting what they see and it is in direct contrast to the viewpoint that truth can be 'discovered' and accumulated piece by piece. The standard method for mapping cognitive constructions is the *Repertory Grid Technique (RGT)*. RGT was originally created by Kelly as a means of understanding variations in how people view phenomena in the world. Variants of the original technique have since been applied in many different kinds of research (see Section 5), based on the notion of RGT as a form of structured interviewing [15].

The research objective for our RGT study was to explore and gain a deeper understanding of the contextual factors that should be considered when selecting or tailoring software development practices to suit specific project contexts. The study was motivated by the now-accepted wisdom that software projects do not follow a single, defined methodology, but rather combine pieces from methodologies in a pragmatic way, often adapting individual practices to suit the project's environment [1, 8, 20, 22, 26, 27, 29, 32, 39]. As we aimed to understand this phenomenon from the perspective of the practitioners and wanted our influence on participants to be minimal, we selected RGT as a suitable approach.

In this paper, we describe the RGT study. We asked twelve practitioners from three software organisations about their viewpoints on the factors affecting practice efficacy. We found that both participants and researchers struggled with several aspects of the RGT technique. We discuss some possible reasons, including our lack of experience with the technique.

Our contributions are the highlighting of the challenges that may occur when implementing this technique, an analysis of the issues encountered and some possible mitigation approaches. These may serve as support for SE researchers considering an RGT based study.

In Section 2, we overview the RGT approach. In Section 3, we describe the approach taken for our RGT study along with details of the study design. In Section 4, we discuss the challenges we encountered as a meta-analysis and in Section 5, we overview related work. In Section 6, we discuss study limitations and in Section 7, we summarise the paper and discuss future work.

2 REPERTORY GRID TECHNIQUE (RGT)

A fundamental postulate of Personal Construct Theory is that a person's actions are channeled by how they anticipates events will unfold [16]. This means that beliefs and understandings held will help shape how future events are approached and effectively influence outcomes. Kelly suggests that the need to confirm and

EASE 2021, June 21–23, 2021, Trondheim, Norway

ACM ISBN 978-1-4503-9053-8/21/06…$15.00
https://doi.org/10.1145/3463274.3463841

disconfirm predictions have greater psychological significance than reinforcements by reward and punishment.

PCT is associated with a number of corollaries. Those most relevant for this study are [16]:

Construction Corollary In order to anticipate how future events will play out, a person creates a mental construction that supports a consideration of similarity and difference. In effect, since no two events are ever identical, replication is not possible.

Individual Corollary People construct events differently — it is unlikely that two individuals will create identical systems.

Dichotomy Corollary "A person's construction system comprises a finite number of dichotomous constructs." A construct is always bipolar.

Experience Corollary A person modifies his system of constructs as he experiences new events. This is associated with increased maturity.

Fragmentation Corollary A person is able to support systems of constructs that are incompatible with one another i.e. the constructs are not necessarily logically consistent.

Commonality Corollary People with similar construct systems have psychologically similar processes.

One key idea emerging from the above is that a person's beliefs can shape future group behaviours, a notion that is increasingly seen as important in the field of Information Systems (IS) [37]. Oher key ideas are that a person's beliefs can change over time, and that people differentiate according to aspects of similarity and difference. The RGT approach thus focuses on comparing instances of the phenomenon under investigation.

The main components of a repertory grid are the *topic*, *elements*, *constructs*, and *ratings* [15, 37]. The *topic* defines the domain of interest and *elements* are the items of interest within this domain. Elements must be homogeneous in nature i.e. should not include a mix of abstract and concrete nouns, or a mix of nouns and verbs. Depending upon the specifics of the study, the topic and elements may be chosen by either the researcher or elicited from interviewees during the study. For example, Young et al. applied RGT to investigate the relationship between personalities and roles in an XP team [41]. As the authors wanted to restrict the study to roles within the organisation (for example, Team Leader, Technical Architect), the roles were chosen by the researchers as the grid elements. For our RGT study, the topic is 'situated software practices'. We did not have specific practices in mind as we wanted to understand the contextual factors that might be generally relevant for practice efficacy. We thus asked study participants to select elements (practices) that were of interest to them. We expected that elements might include practices such as 'design review' or 'quality sign-off'.

Constructs capture characteristics of elements that manifest as bipolar descriptions. Poles may represent either *opposites* or *contrasts*. Constructs serve to establish a person's viewpoints with respect to the elements and are "formed by the individual's personal interpretations of the issues associated with the element" [41]. In the IS domain, these are generally elicited from interviewees during the RGT process. For construct identification, the recommended approach is to ask the participant to consider elements in sets of three (triads) and, for each set, to identify in what way the first two are similar and the third different. In this way, the two poles of a construct emerge. This construct represents the perspective of the participant, unsullied by the opinions of the researcher. In the Young et al. study, two constructs that emerged were 'Diplomatic with people - Strong analytic ability' and 'Flexible — More individualistic' [41]. In both cases, the poles describe contrasts rather than opposites. Constructs that emerged during our RGT study included 'Learned how — Learned why' and 'We had the relevant information - We did not have the information we needed' (see Figures 1 and 2). For the element 'design sprint', an example of a construct might be 'Team collaborated well - Disfunctional team'.

According to Kelly, a person's belief system comprises a hierarchy of concepts, characterised by cause-and-effect ('implies') relationships and with more abstract concepts at the top i.e. 'implied by' lower level concepts. This structure was explored by Hinkle, a Ph.D. student of Kelly, who wished to understand resistance to change in an individual's personal constructs [14]. Hinkle perceived that a construct must be understood within its hierarchical context and proposed a method, 'laddering up', for eliciting super-ordinate constructs i.e. the participants core values. Concepts nearer to the top of the hierarchy can be elicited by asking questions such as 'Why is that important to you?'. For example, Koreni describes a laddering proccess where an initial construct of "Wide (narrow) area of expertise" 'laddered up' to "Not expendable (expendable)" and then to "Financial stability (insecurity)" [21]. Later researchers extended the approach by including 'laddering down' to help participants clarify meaning. This involves asking questions such as 'In what way are the bipolar values different?'.

Ratings enable people to express a viewpoint on the extent to which a charactersitic is represented and are used when researchers want to compare or amalgamate the viewpoints of several individuals. This is done by assigning a scale to the construct, with the ends of the scale representing the poles of the construct. For example, for the described construct for the 'design sprint', a five point scale would allow a participant to state their view on how the characteristic was manifested for the element. Participants in the Young et al. study rated constructs according to how they perceived the importance to each role of the personality characteristic described by the construct. Ratings were then amalgamted to identify which characteristics were most highly associated with each role [41]. We did not include ratings in our RGT study as we did not want to compare or amalgamate information.

The research objective thus defines the specific technique to be implemented, in particular, whether the elements and constructs are determined by the researcher or the participant. In general, when the goal is to identify emerging themes, participants select both elements and constructs, whereas elements are selected by the researcher when the goal is to compare and/or analyse.

3 REPERTORY GRID STUDY

3.1 Research Motivation

The research objective for this RGT study is to *understand practitioners' viewpoints on the contextual factors that affect the effectiveness of situated software practices*. The motivation for this is presented below.

The history of software engineering process has been one of *advocacy*. Many authors, from both academia and industry, have architected software development methodologies and processes in the belief that strict adherence by practitioners will inevitably result in project success. Examples include traditional models such as waterfall [34] and spiral [3] and agile methods such as eXtreme Programming (XP) [2] and Feature Driven Development (FDD) [25]. The wisdom manifested was that "development models are best regarded as a coherent set of practices, some of which are required to balance the performance trade-offs arising from the use (or absence) of others" [7].

As time has passed, it has become clear that software projects simply do not follow a single, defined method to the letter, but rather adapt practices to suit the project's environment [1, 8, 20, 22, 26, 29, 32, 39]. The notion of 'tailoring according to contexts' has become the new wisdom.

In parallel with the move towards ad hoc tailoring, those involved in software process improvement (SPI) have experienced a shift away from a formal, pre-planned set of activities towards a *continuous learning* paradigm where the aim is to "gradually change SPI into a more project-integrated activity" with adaptation often happening "in response to customer demands" [22]. The trend is for SPI to become connected to "specific project characteristics" with the result that "transfer to the organization level is often not reasonable".

These paradigm shifts raise two issues. First, we cannot know exactly what a practice entails without an awareness of the context in which it was enacted i.e. we must now view practice and context as inherently intertwined. This notion has precedence in the organisational and management sciences [30] and indeed our previous research has exposed this dependence for software practices [18]. Second, we must now be concerned about the fact that practitioners may be adapting practices in a way that reduces effectiveness. Legnick-Hall and Griffith suggest that applying knowledge in an intuitive or experimental way introduces a lack of fit between type of knowledge and how it is applied. A practice is generally expected to achieve a known outcome i.e. is a 'strategic asset' whereas the notion of 'tinkering' is appropriate only for innovation and creative activities [23]. From this perspective, ad-hoc tailoring that is not grounded in evidence might be viewed as a 'hidden' issue.

Our previous explorations have led us to adopt the position that, before we can confidently support practitioners in their tailoring efforts, we must first build an evidence base that relates a situated practice (a practice performed within a specific context) and its outcomes with respect to desired objectives. A prerequisite for this is a deeper understanding of *context* - without a suitable, agreed framing of context, we cannot progress with evidence accumulation and so cannot be in a position to advise confidently, in an evidence-based way. We thus view the abstracting of context as a crucial step in the domain of software process research.

Several researchers have introduced frameworks for software context [1, 6, 33]. We have critiqued these in earlier works as lacking in consistency and completeness and have proposed an abstraction that aims to address these issues [19]. However, before carrying out a formal evaluation and refinement study, we wanted to elicit viewpoints from practitioners that were not influenced by our own biases.

3.2 Research Method Selection

Stol and Fitzgerald critique the literature on research methodologies within the software engineering community, in that these tend to "enumerate a set of methods, without a holistic view that affords a systematic comparison" [36]. The authors posit the need for a more holistic approach and propose a research strategy framework, adapted from earlier work in the social sciences. The framework focuses at the level of the *research strategy* and is characterised by two dimensions, the level of *obtrusiveness* within a given setting and the degree of *generalisability* to other settings [36]. Eight archetypal strategies are identified. Strategies highest on the 'obtrusiveness' dimension are *Laboratory experiments* and *Experimental simulations* as these involve highly contrived settings. Strategies low on 'generalisability' include *Field experiments* and *Field studies* as results tend to be specific to the studied environment i.e. the study takes place in natural settings. A *field experiment* is more obtrusive than a *field study* as it involves a greater manipulation of the research environment, for example, as in an action research study [36].

Our aim of eliciting a deep understanding of practitioners' viewpoints prompted the decision to base information gathering on one-on-one interviews. Responses will be "in relation to some stimulus (question or instruction), independent of setting" [36]. Based on the framework, our research approach maps to a *Judgement study*. The choice of setting is not relevant i.e. neutral setting, The researcher aims to elicit information using some kind of structured approach (reasonable level of obtrusiveness) and wishes to generalise based on responses. As we wanted a highly structured approach with minimal inputs from the researchers, we implemented a repertory grid study.

3.3 Repertory Grid Design

Edwards et al. describe a set of design decisions that must be made to ensure a) grid reliability and validity and b) the suitability of the grid for the research questions [10]. We discuss these in relation to our RGT study.

Our aim was to elicit opinions and expertise about situated software practices with minimal inputs from researchers. We believed that asking participants to choose practices would result in their selecting practices that were of greater interest to each practitioner [37]. According to Edwards et al., this would result in a richer set of constructs [10]. We thus chose to supply the topic (*Situated software practices*) and participants chose both elements (*practices*) and constructs (*contexts*). We guided selection by asking participants to include practices they perceived as being successful and those perceived as unsuccessful [37]. As we did not want to compare or analyse responses concerning practices, we did not ask participants to rate practices.

For elicitation, we planned to apply a triadic approach with six elements, as recommended in the literature [10]. However, we did have a concern that the elicitation process might be too long and tested this in a trial. As we wished to elicit a rich set of constructs rather than looking for some kind of consensus, we interviewed participants individually. The plan was to keep the interview time within one hour, as we understand that practitioners are busy and we believed that we should constrain time taken. For construct poles, we aimed for *contrasts* (for example, ‘publication speed’ — ‘review rigour’), as opposed to *opposites*, as this is claimed to give a richer set of data [10].

In a bid to avoid responses affected by “social desirability bias” i.e. where participants may “provide responses that are ... in line with the impression they want to create” [17], we explained that there were ‘no correct answers’ and that information given would not be reported back to management. When eliciting constructs, we applied the guidance offered by Jankowicz [15]. When asking participants to select two ‘successful’ and two ‘unsuccessful’ practices, we provided alternate terms i.e. the practice ‘was successful / unsuccessful’, ‘went well / badly’, ‘was effective / ineffective’ and ‘was liked / disliked’. The aim was to avoid researcher influence. To aid the participant, we set down the selected elements (practice instances) on cards. For each construct offered by a participant, we a) worked with the participant to establish the intended meaning (for example, ‘happy / sad’ versus ‘happy/angry’), and b) asked ‘in what way?’ and/or ‘what do you mean?’ in order to clarify concepts (laddering down). We also applied a laddering up process to identify the super-ordinate contructs by asking ‘Why did this matter?’. Higher-level constructs represent core values i.e. the essential elements that make a practice successful. Remaining nodes represent contexts at different levels of granularity. Our intention was to combine threads for a participant whenever there was a node in common as this would provide insight into the participant’s core values and the main context perceived as affecting these. The approach is variation of that implemented by Pankratz and Basten [31].

We implemented a trial with personal contacts actively involved in software engineering activities. This exposed some issues. When asked about ‘things you do’, one participant described some practices in terms of events. We thus had difficulty in reconciling the two notions of ‘event’ and ‘practice’ and establishing ‘similarities’ and ‘differences’. The participant also tended to describe the practices rather than focusing on contrasts. We also found that including six elements was too time-consuming. As a result, we introduced a general discussion with the participant about ‘practices you regularly carry out’ before asking them to identify specific practice instances and we took care to emphasise that we would focus of similarities and differences rather than description. We also asked the participant to choose four elements instead of six.

The steps we used for information gathering are summarised as:

(1) We chose the topic in advance and asked the participant to discuss in a general way practices regularly carried out as part of their role.
(2) We asked the participant to select two ‘successful’ and two ‘unsuccessful’ practices that they had recently been involved in. We explained that, by ‘practice’, we did not mean only the ‘standard’ practices, for example, ‘design review’ but any activity that the participant believed to be interesting for achieving outcomes. To support the process, we labelled the ‘successful’ practices A and B, and the ‘unsuccessful’ practices C and D.
(3) We asked the participant to focus on the two ‘successful’ (A and B) and one ‘unsuccessful’ (C) practice. We asked ‘In what way are A and B similar, and different from C?’ We assured the participant there was no ‘correct’ answer. When happy that we understood the two poles of the construct, we noted these with the similar pole on the left.
(4) We discussed with the participant to ensure we understood what contrast was being expressed, modifying appropriately.
(5) We then commenced a process of laddering by asking the participant ‘Why did this matter?’ or ‘What was the result of this?’ (laddering up) and ‘Can you be more specific?’ (laddering down). There was no pre-defined ordering for these questions, rather we tried to adapt to the thinking processes of the participant. As each element was defined, we noted these, linking to the existing nodes to form a thread.
(6) We repeated steps 4-6 for ACD, ABD and BCD.

Table 1: RGT study participants.

Role	Org	Role
Product Manager	Small	>5 yrs
Business Analyst	Large	>10 yrs
Product Owner	Large	<5 yrs
Software Architect	Large	<5 yrs
UX Architect	Large	>10 yrs
Developer	Large	<5 yrs
Developer	Large	>5 yrs
Developer	Small	>5 yrs
Software Engineer	Large	? yrs
QA Analyst	Large	<5 yrs
Test Analyst	Large	>5 yrs
Specialist Tester	Large	? yrs

After the first few interviews, we found participants were experiencing difficulties with element selection. We amended the process to include two practices only, with participants choosing one ‘successful’ and one ‘unsuccessful’ instance for each. We discuss in Section 4.1. Interviews were attended by one or both of the authors and audio recorded. The first author transcribed the audio, and checks were carried out on each transcript by the second author. We implemented a number of standard ethics research protocols, for example, relating to permission, anonymity and confidentiality.

3.4 Participants

As our intention is to include information in a general model, our approach is essentially one of building theory from cases. According to Eisenhardt and Graebner, *theoretical* as opposed to *random* sampling is appropriate [11]. The organisations approached for participation were selected from a list of organisations known to the authors. However, the organisations varied in size and application area and the participants represented a variety of software engineering roles. In Table 1, we overview the participants. We discuss the missing ‘Role’ data in Section 4.1.

The main contact for each organisation, the *manager*, was a senior manager with authority to make decisions relating to the software development function. The names of individuals from within organisations were provided by the manager after a preliminary meeting to discuss the study. We believed the manager would at this point have a clear understanding of what we hoped to achieve and would be more likely to provide a realistic cross-section of participants. Talking to employees other than those provided by management was not considered both for formal ethical reasons and because we would view this as a serious breach of trust. We asked the manager to provide participants covering a range of roles and backgrounds.

3.5 RGT Study Evaluation

The notion of evaluating measurement instruments and protocols originated in the concrete, physical world, where instruments included equipment to measure, for example, weights of objects, quantity of current flowing along a wire, etc. The key concepts were *reliability* (do multiple measurements centre around the same value) and *validity* (does this value accurately represent the quantity being measured). In this domain, objectivity is of primary concern. As research artifacts and processes in experimental research are considered to be measuring instruments and protocols, the same notions have been applied in this area. For research based on a positivist philosophy, where the belief is that knowledge is based on "logical inference from a set of basic, observable facts" [9] and objectivity is the goal, the notions of reliability and validity continue to be appropriate. However, this is not obviously the case in the domain of qualitative research, which is contextual and subjective rather than generalisable and objective. A consequence has been a long and heated dialogue about what reliability and validity actually *mean* for those involved in qualitative research [4, 24, 28]. According to Whittlemore et al., attempts to adhere to the original criteria have resulted in studies exemplified by "procedural charade and pseudoscience" [40]. Edwards et al. point out that the application of the traditional measures to RGT studies can be problematic because a) the items to be measured emerge during the research, and b) grids are used in many ways and concepts of reliability and validity will vary according to the type of grid [10].

Despite the above issues, Whittlemore et al. suggest that validity is "an acccurate term" for which appropriate criteria can be developed. The authors propose a "synthesis of contemporary viewpoints" [40], which comprises primary and secondary criteria. We have adopted this scheme and evaluate our RGT study as below.

The primary criteria are considered to be necessary for all qualitative studies. The first two, *credibility* and *authenticity* relate to validity threats of bias and distortion of participant meaning. *Criticality* and *integrity* relate to a lack of attention to the applicability of the research process, to unexpected findings or alternative perspectives on outcomes.

Credibility refers to the level of confidence that the data has been collected and interpreted accurately. For elicitation of information, we selected an RGT variation that involves participants selecting both elements and constructs (minimal researcher influence) and includes upwards and downwards laddering (deeply understand the participant's perspective). We did not carry out any deep analysis of the data.

Authenticity relates to adherence to the participants' meanings and perception. We see two possible risks. The first relates to possible distortion of meaning by the researchers. This was addressed by the laddering down process, as this is aimed at ensuring a deep understanding of meaning. A second risk is that of the participant providing data that is "socially acceptable or ... in line with the impression they want to create" [17]. We addressed this by a) explaining to participants that there was no 'right answer' and that responses would not be reported back to management, and b) including both upwards and downwards laddering in the RGT procedure, as we believed that the need to 'explain' lower and higher constructs would expose any lack of basis in reality.

Criticality requires that the researcher assess both process and results, for example, by critiquing the process, exploring negative findings and examining biases. For our study, the concept of 'negative findings' does not apply. There was, however, a possible bias in that the first author has been involved in model-building for context and may have been biased towards 'finding' constructs that supported their beliefs. We addressed this by a) ensuring one of the interviewers was not involved in earlier model creation, and b) strictly adhering to the process. However, we have reservations about the use of RGT for this kind of study, and discuss this in section 4.3.

Integrity relates to qualitative research, where researcher subjectivity inevitably affects data analysis. Integrity may assured by following process during implementation and analysis. Although we followed the RGT process, in this paper, we do not analyse the RGT data.

Secondary criteria are "additional guiding principles" to be applied as appropriate [40].

Explicitness refers to the existence of an audit trail in order that the investigation can be followed by others. We have recorded all data and decisions made.

Vividness concerns the presentation of rich data and is not relevant for our study.

Creativity relates to the use of novel methodological designs. We are following a standard process and so the criterion is not relevant for our research.

Thoroughness requires that data collection proceeds until no more useful information is being elicited and that themes are fully developed. We did not address this criterion.

Congruence relates to theoretical connectedness throughout the study. We applied an accepted approach in a consistent way and believe the criterion is met.

Sensitivity refers to a respect for the participants' needs. The ethical considerations and procedure described in section 3.3 were architected with these needs in mind.

4 META-STUDY ANALYSIS

In this Section, we present as a meta-study the challenges we met during RGT implementation, suggest some possible mitigations, and discuss our findings. To support the cross-referencing of challenges to mitigations (see Table 3), we give each challenge a descriptive label. For example, the label for the tendency to describe practices in a general way is 'PRACTGEN'.

Table 2: Selecting practice instances.

I wonder if I should get my computer to help me remember.
... go and get my computer
... have a look at the jobs that we've recently done.
It's going to be an awkward one. I hate doing the same thing
I try and do ... different things or different ways ...
I want to ... my laptop and look at the ones I've done.

4.1 Challenges

As the challenges were exposed during the RGT process, we did not have in advance a meta-study protocol to use when conducting interviews. Our approach was to provide empirical data to support our experiences from a post-analysis of the interview transcripts. We carried out thematic analysis and applied selective coding to extract data that related to the issues we encountered [5]. Analysis was done by the first author and themes discussed and confirmed at regular meetings with the second author. The themes that emerged were *Time Delays*, *Data Integrity* and *Causal Pathways*. Below, we describe each of the themes along with some supporting evidence from the transcripts.

4.1.1 Time Delays. [PRACTGEN] There was a tendency for some participants to want to describe chosen practices in general (*"I can think of the general situation"*). During interviews, we had to continually bring focus back to specific practice instances and to comparison rather than description (*"That was the result of the practice, though, rather than the instance?"*). This was frustrating for both interviewer and participant. As we did not want to cut the participant short when describing a practice, and did not want to go over the hour interview time, we found we had less time available for exploring instances. The concern around time led to a failure to ask two participants about length of time in the role (see Table 1).

[INSTSEL] Some participants experienced difficulty in identifying practice instances. In some cases, the reason was uncertainty about what a practice is (*"the first stage is quite interesting"*). This represents a limitation of our research (see Section 6). In other cases, it simply appeared to be difficult for participants to remember specific occasions (*"the whole thing is working quite well"*, *"trying to think about something very specific"*, *"I'm just trying to think of some very specific pull requests"*). In such cases, we found that a large part of the interview was spent on discussions around practices and practice instances and this reduced the time we had available for identifying similarities and differences. As this demands significant effort to ensure constructs are really understood, we had a concern that in some cases findings would be superficial. An analysis of the transcripts showed that, in five of the twelve interviews, these initial discussions lasted longer than fifteen minutes. In one case, we did not begin examining triads until 26 minutes into the interview. In Table 2, we illustrate the problem with some further participant comments.

4.1.2 Data Integrity. [INSTPART] In a small number of cases, the practitioner addressed the issue of being unable to identify suitable practice instances by describing an instance they had not participated in (*"there was a point where ... I was no longer directly involved"*, *"I was doing archeology on that one, so I don't know the process at the time"*, *"that's what I imagined happened."*). This was not clear to the researchers until well into the body of the interview. The identification of causal links became problematic because the participant, although knowledgable about 'the problem', was not in a position to discuss the cause(s) of the problem.

[CONSTRIDEN] In the earlier interviews, we allowed participants to identify different practices i.e. potentially four different practices. However, we found that often the practices selected were so different that participants struggled to find similarities and differences i.e. to identify constructs (*"all the major things I thought about wasn't about comparison. It was more like preferences"*, *"I'm trying to get what the actual contrast is"*, *"Ah it's tricky"*). This caused us to change strategy and ask for one 'good' and one 'bad' instance for each of two distinct practices. However, the difficulties in identifying constructs remained, with the associated uncertainty about data integrity.

4.1.3 Causal Pathways. [LADDERISS] Another set of construct-related challenges concerns the laddering process, in which causes (laddering down) and effects (laddering up) are identified. In a small number of cases, the initial construct identified described the outcome of the practice, for example, 'found many errors - found few errors'. It then seemed really difficult to identify the 'causal' contextual factors. Generally, identifying causes and effects was problematic. For some participants, this may have been because the effect was 'obvious', for example, 'late delivery led to unhappy customer'. For others, the participant produced a rich set of 'causes' or 'effects' and the challenge became one of disentangling the threads. This meant that what was effectively the same question was being asked over and over, in a slightly different way. This appeared to be frustrating for some. In some cases the interviewer had to prompt when the participant was having diffculty. This had the result of effectively influencing the participant, a situation the RGT approach is designed to avoid.

[THREADCOMPL] We found that participants varied enormously in the number and complexity of threads emerging from a single triad. For example, in Figure 1, we show one of the more simple outcomes. The discussion started at the centre node. Laddering down exposed root cause (relating to personalities) and laddering up clarified the outcome (effort required for understanding).

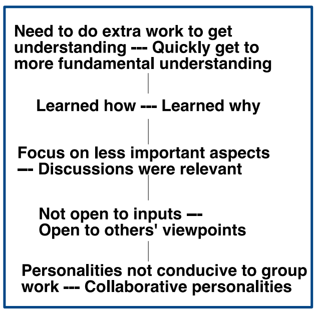

Figure 1: RGT - simple thread

In Figure 2, we show a more complex outcome. In this case, the initial construct was quickly and easily identified. However, during

Table 3: Issues and mitigation approaches

Issue	Mitigation
PRACTGEN	Clarify focus at start
INSTSEL	Ask to identify before interview
INSTPART	Clarify at start
CONSTRIDEN	Choose same practice in all cases
LADDERISS	Explain that 'obvious' is ok
THREADCOMPL	No mitigation.

laddering, a rich body of information was provided in response to each laddering prompt, each aspect of which was highly relevant (*there's two outcomes, really*). Focusing the discussion to pursue a single thread in depth was time consuming and frustrating. It appeared that many of the concepts were highly entwined in the participant's mind and simple causation almost impossible to ascertain. For example, the 'causes' of good collaboration and team satisfaction may have included any of the nodes within the large rectangle.

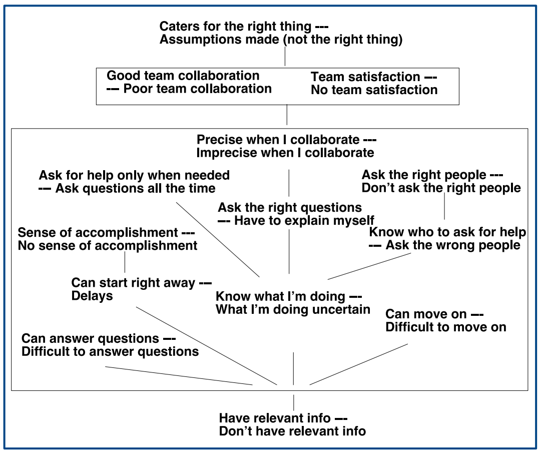

Figure 2: RGT - complex threads

4.2 Mitigation Approaches

As reseachers new to RGT, we next considered some mitigation approaches for the issues discovered and that we might implement in followup studies. In Table 3, we summarise the issues along with our thoughts on addressing these.

One approach to addressing the diffculties experienced in identifying specific practice instances might be to ask participants to prepare these in advance. We do not know how effective this might be, as practitioners are busy and would possibly not be in a position to make this preparatory effort.

The issues of describing the practice in a general way and selecting instances not participated in can be addressed by spending more time to clarify expectations at the beginning of the interview. However, the RGT approach is already time-intensive and our belief is that more time up-front would mean that the discovery of similarities and differences would become a superficial exercise.

We are unable to present a mitigation approach for the challenges with identifying causes and effects. One factor contributing to this phenomenon might be that people are not often required to introspect on what they are doing and so the interview represented something quite new for them. The *Fragmentation Corollary* suggest that systems of causal networks are not necessarily logically consistent and so it might be that it is inherently difficult for people to discuss causes and effects in a meaningful way. A third factor might be that the development environment is too complex for specifying simple causation at such a granular level. Pinning down 'what caused what' when discussing team functioning, task characteristics, outcomes and team member reactions is possibly an impossible ask.

Problem space complexity might be reduced by constraining the participant base to include those who have carried out the same practice in different teams within the same company. This would mitigate for some of the issues in Table 3. However, it would not address the time issue and it is possible that RGT is not the best approach to use in such a complex environment.

4.3 Discussion

Our experiences with RGT lead us to question the 'fit' between a technique that is based on the notion of 'no two people think alike' and the desire to create a research framework that effectively represents a perspective on 'what is really there'. When applying RGT, we are effectively assuming there is a 'fixed' set of constructs which relate in a causal way and which can be exposed by amalgamating the mental constructs of individuals. However, this is inconsistent with many of the RGT Corollaries (see Section 2).

Kelly's original notion implies that an individual perceives different instances of events in a different way (*Construction Corollary*) and this is reflected in the mental constructs they build or alter as a result of the event. The *Individual Corollary* implies that it is unlikely that different people will view the same event in the same way, and the *Fragmentation Corollary* implies that a person may hold conflicting mental constructs. These are all inconsistent with the notion of a 'fixed' external reality.

The observation that mental constructs describe the individual rather than the event caused us to consider whether insights into the personalities of the participants might be inferred from aspects of their grids. In particular:

- similarities (*Commonality Corollary* and variation (*Individual Corollary* in the 'richness' of the threads provided by different participants (see Section 2)
- the number of, and variation in, key themes between triads for a given participant
- the concepts appearing at the top of the threads after laddering up, as this is associated with core values [14].

We are exploring these questions in parallel research.

Our planned strategy of identifying core elements of software development context by means of laddering and amalgamation of threads did not have the results expected. However, we did elicit a rich set of information about context. In a parallel investigation, we are applying thematic / content analysis to analyse this data, perhaps by identifying themes common to several participants. This study is in progress and at this point and we cannot formally discuss findings.

However, we believe that implementing an RGT study was not the best way to elicit this data as it was too time consuming. In addition, the contexts exposed were focused on people and constructs describing objective context were minimal. This is illustrated in Figures 1 and 2, where the majority of nodes relate to personality and team related concerns. We found that factors such as product characteristics, for example, 'product type', locational issues, for example, 'different time zones', and stakeholder constraints, for example, 'client delivery expectations' were not often included in the causal networks. This means we were successful in eliciting only a subset of the kinds of information we were hoping to obtain.

In our experience, the technique was demanding on both interviewer and interviewee. We acknowledge that our inexperience with the technique may have been a significant contributing factor to the problematic implementation. Another possible factor is the lack of fit between a methodology that is targeted at understanding peoples' perspectives on a topic and an objective that concerns capturing some external reality. It may also be that the RGT design decisions must be highly tailored to the specific research objective (see Secton 5.2). These represent rich areas for future study.

5 RELATED WORK

In this section, we describe studies within the computing domain that have applied RGT. We examine how they applied RGT. We also compare and contrast our RGT study with those reported in the literature.

5.1 Application of RGT

Two survey papers examined RGT studies in empirical software engineering [10] and Information Systems (IS) [37]. Edwards et al. point out that RGT sits with interpretive research rather than with positivist approaches, i.e. the "focus is on understanding, before developing theories that can be subsequently proved (or disproved)" [10]. Both papers conclude that RGT can be a useful technique in the computing domain. Edwards et al. provide some recommendations to researchers employing RGT in empirical software engineering, including ensuring clear research questions are defined [10]. However, neither study describes explicitly the challenges in applying an RGT approach.

Edwards et al. identified eight studies that apply RGT in the field of Software Engineering [10]. Tan and Hunter describe two additional studies in the Information Systems field [37]. Topics covered by the studies found in these survey papers relate mainly to human aspects and include characteristics of team members, team traits and roles, training, quality of systems analysts, and system interface design. Other studies address project risks, software process improvement, requirements, and rule development for expert systems. In addition to the papers described in the two survey papers, additional studies have used the RGT approach in the computing domain [13, 31, 35, 38, 41]. Again, the majority of these studies have focused on human aspects including studies on the relationship between personalities and roles in an XP team [41], situational factors that influence team performance on successful projects [35], and managers' perceptions of IS project success mechanisms [31]. Two of the studies were more technical in nature and focused on tacit architectural knowledge [38] and attributes perceived as being indicative of quality in IT projects [13] (though human aspects were also identified as attributes in this study).

For the studies described in the two survey papers, most use partial or fixed repertory grids, where the elements and sometimes even the constructs are provided by the researchers. Full grids, where the participants provide both the elements and constructs, as was done in our RGT study, were used in only three of the ten studies. However, Edwards at al. argue that one of these was not actually using a full repertory grid as described. Edwards et al. suggest that the full repertory grid approach would be suitable for studies focused on behavioural or organisational aspects [10]. The studies we found applying RGT in the computing domain outside of these two survey papers mostly use a full repertory grid [13, 31, 35, 38], with only one using a partial repertory grid [41].

The above studies suggest that an RGT approach has been applied to address a range of topics in the computing domain, though most are related to human aspects. Most use a partial or fixed repertory grid.

5.2 Comparison with Related Work

In this Section, we compare and contrast our RGT study with the other RGT studies in the computing domain with the aim of shedding light on some of the challenges we faced.

In our RGT study, we used a full repertory grid. Participants supplied the elements (software practices), since we did not know in advance which practices the participants were using. Interestingly, for the six studies that used full repertory grid, five of the six used software projects or software systems as the elements [10, 13, 31, 35]. The other study that used a full repertory grid used architectural choices as the elements [38]. It is possible that a 'software project', 'software system', or 'architectural choice' is more concrete than a 'software practice', making it easier to identify instances. An instance of a software project, for example, includes all activities associated with a single software project. On the other hand, a single instance of a software practice could be as small as a single meeting (e.g. a daily stand up). It could have been difficult for participants to remember details of specific instances of software practices. This could also be a reason for achieving less variety in the constructs as was achieved in prior studies and potentially a reason laddering was difficult. Participants had more experiences to draw on in identifying constructs at the level of a project compared to a single instance of a practice. We suggest researchers consider these aspects when designing a RGT study to ensure the elements are at an appropriate granularity.

We used a triadic elicitation approach, where participants compared three items at a time. They looked for reasons why two of the three were similar that was different from the third element. Edwards et al. [10] reported that a triadic approach is easier for participants compared to examining dyads. A triadic approach is also in line with Kelly's Personal Construct Theory [16]. However, Fransella et al. has argued that triads can be cognitively challenging, and that a diadic approach could be easier [12]. We note that some of the other studies used a diadic approach [13, 31]. It may be that comparing two elements (diadic approach) is inherently simpler than comparing three (triadic approach). Perhaps given the complexity of software practices, a diadic approach would have been better suited. We suggest researchers looking to apply RGT trial both elicitation approaches to identify the one that is best suited to the topic under exploration.

In the study by Tofan et al. [38], one participant is reported as commenting that the session proved "difficult, because you have to think so much". Participants also reported disliking the time consuming nature of RGT. This finding is consistent with the frustrations experienced during our study. The other papers that reported on studies that applied RGT in the computing domain did not explicitly describe challenges or difficulties. However, this does not mean that similar challenges to those described in this paper were not experienced. We suggest researchers who apply RGT consider following the precendent of Tofan et al. and explictly report challenges (and mitigation strategies) so that future researchers can learn how to best apply RGT [38].

6 LIMITATIONS

The main limitation for this RGT study is the researchers' lack of experience with RGT. According to the literature, RGT implementation requires experience and, for both researchers, this study was the first in which RGT had been the research approach selected.

One result of this lack of experience was that we did not preempt the difficulty participants would have in identifying elements (practice instances). Our trial did expose some possible issues and we addressed these by introducing a conversation around practices at the start of interviews. This mitigation failed to remove the problem. It is possible that the reason is that we did not define 'practice' but rather used terms such as "something you do regularly as part of your job" and "any atomic thing, like a design review" and "meeting with the customer to try and pin down requirements". However, the problems we encountered did not appear to be consistent across cases and another contributing factor may have been the fit between technique and problem characteristics.

A limitation of the research is that we report on issues with the research approach taken when the research objective was to implement and report on a study to elicit information on software practices context. This means that the findings reported in this paper relate to a meta-study of the RGT approach when much of the description relates to the planned RGT study. Although a serious limitation, we believe our contribution to be of value to those planning an RGT based study.

7 SUMMARY AND FUTURE WORK

In this paper, we describe an RGT study to explore and gain a deeper understanding of the contextual factors that should be considered when selecting or tailoring software development practices to suit specific project contexts. As we did not want to influence the outcomes, we selected the Repertory Grid Techique (RGT) as a suitable approach. This technique is claimed to be suitable for understanding peoples' perspectives on a topic at a deep level. We selected the topic as 'software practices' and asked participants to choose elements (an instance of a software practice implementation) and constructs (the contextual similarities and differences between instances). We employed laddering to create causal threads that manifested a participant's beliefs concerning the success of the practice. We hoped to expose key contexts by amalgamating the participants' threads.

We encountered many challenges during RGT interviews. Participants found identifying specific practice instances difficult, tended to describe a practice in general rather than focusing on similarities and struggled with clarifying causal links. This meant that some interviews were frustrating for both researchers and participant.

We found that each identified challenge manifested in a sub-set of the interviews. This led us to posit that the success of the RGT technique may be context-dependent and may not represent the best approach for understanding complex situations involving team functioning, task characteristics, task outcomes and team member viewpoints. Our contribution is to add to the available literature on RGT in SE by highlighting and analysing possible challenges. This may serve as support for researchers considering an RGT based study.

We did elicit a large set of data from the interviews and are in the process of analysing this. We have identified two areas for future work. We will investigate the possibility of gaining insights into personalities from grid aspects. We will also explore themes for software context in situated software practices.

REFERENCES

[1] David Avison and Jan Pries-Heje. 2008. Flexible Information Systems Development: Designing an Appropriate Methodology for Different Situations. In *Enterprise information systems : 9th International Conference, ICEIS 2007*, J. Filipe, J. Cordeiro, and J. Cardoso (Eds.). Springer, Berlin, Heidelberg, 212–224.

[2] Kent Beck. 2000. *eXtreme Programming eXplained - Embrace Change.* Addison-Wesley, United States of America.

[3] Barry W. Boehm. 1988. A Spiral Model of Software Development and Enhancement. *IEEE Computer* May, 11 (1988), 61–71.

[4] Denny Borsboom and Keith A. Markus. 2013. Truth and Evidence in Validity Theory. 50, 1 (03 2013), 110–114.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Paul Clarke and Rory V. O'Connor. 2012. The situational factors that affect the software development process: Towards a comprehensive reference framework. *Information and Software Technology* 54 (2012), 433–447.

[7] Michael Cusumano, Alan MacCormack, Chris Kemerer, and Bill Crandall. 2003. Software Development Worldwide: The State of the Practice. *IEEE Software* 20, 6 (Nov/Dec 2003), 28–34.

[8] Mariana de Azevedo Santos, Paulo Henrique de Souza Bermejo, Marcelo Silva de Oliveira, and Adriano Olimpio Tonelli. 2011. Agile Practices: An Assessment of Perception of Value of Professionals on the Quality Criteria in Performance of Projects. *Journal of Software Engineering and Applications* 4 (2011), 700–709.

[9] S. Easterbrook, J. Singer, MA. Storey, and D. Damian. 2008. Selecting empirical methods for software engineering research. In *Guide to Advanced Empirical Software Engineering*, F. Shull and J. Singer and D.I.K Sjøberg (Ed.). Springer International Publishing, London, UK, 285–311.

[10] Helen M. Edwards, Sharon McDonald, and S. Michelle Young. 2009. The repertory grid technique: Its place in empirical software engineering research. *Information and Software Technology* 51 (2009), 785–798.

[11] Kathleen M. Eisenhardt and Melissa E. Graebner. 2007. Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal* 50, 1 (2007), 25–32.

[12] Fay Fransella, Richard Bell, and Don Bannister. 2004. *A manual for repertory grid technique.* John Wiley & Sons.

[13] Joana G. Geraldi, Elmar Kutsch, and Neil Turner. 2011. Towards a conceptualisation of quality in information technology projects. *International Journal of Project Management* 29 (2011), 557–567.

[14] Dennis Neil Hinkle. 1965. *The Change of Personal Constructs from the Viewpoint of a Theory of Construct Implications.* Ph.D. Dissertation. The Ohio State University.

[15] Devi Jankowicz. 2004. *The Easy Guide to Repertory Grids.* John Wiley and Sons, Ltd.

[16] George A. Kelly. 2017. A brief introduction to personal construct theory. *Consntruttivismi* 4 (2017), 3–25. https://doi.org/10.23826/2017.01.003.025

[17] Carole L. Kimberlin and Almut G. Winterstein. 2008. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy* 65 (2008), 2276–2284.

[18] Diana Kirk and Stephen G. MacDonell. 2016. An Ontological Analysis of a Proposed Theory for Software Development. In *Software Technologies - ICSOFT 2015*, P. Lorenz et al. (Ed.). Communications in Computer and Information Science (CCIS), Vol. 586. Springer International Publishing, Switzerland, 1–17. https://doi.org/10.1007/978-3-319-30142-6_9

[19] Diana Kirk and Stephen G. MacDonell. 2018. Evolving a Model for Software Process Context: An Exploratory Study. In *Proceedings of the 13th International Conference on Software Technologies (ICSOFT 18)*. SCITEPRESS, Porto, Portugal, 296–303.

[20] Jil Klünder, Dzejlana Karajic, Paolo Tell, Oliver Karras, Christian Münkel, Jürgen Münch, Stephen G. MacDonell, Regina Hebig, and Marco Kuhrmann. 2020. Determining Context Factors for Hybrid Development with Trained Models. In *ICSSP '20, Proceedings of the 2013 International Conference on Software and System Processes*. Association for Computing Machinery, 61–70. https://doi.org/10.1145/3379177.3388898

[21] Bojan Korenini. 2012. Conducting Consistent Laddering Interviews Using CLAD. *Advances in Methodology and Statistics* 9, 2 (2012), 155–174.

[22] Marco Kuhrmann and Jürgen Münch. 2019. SPI is Dead, isn't it? Clear the Stage for Continuous Learning!. In *ICSSP '19, Proceedings of the 2019 International Conference on Software and System Processes*. Association for Computing Machinery, Montréal, Canada, 9–13.

[23] Cynthia A. Lengnick-Hall and Robert J. Griffith. 2011. Evidence-based versus tinkerable knowledge as strategic assets: A new perspective on the interplay between innovation and application. *Journal of Engineering and Technology Management* 28 (2011), 147–167.

[24] Robert W. Lissitz and Karen Samuelson. 2007. Further Clarification Regarding Validity and Education. *Educational Researcher* 36, 8 (2007), 482–484.

[25] Jeff De Luca. n.d.. Agile Software Development using Feature Driven Development (FDD). http://www.nebulon.com/fdd/

[26] A. MacCormack, W. Crandall, P. Henderson, and P. Toft. 2012. Do you need a new product-development strategy? *Research Technology Management* 55, 1 (2012), 34–43.

[27] Zainab Masood, Rashina Hoda, and Kelly Blincoe. 2020. Real World Scrum A Grounded Theory of Variations in Practice. *IEEE Transactions on Software Engineering* (2020).

[28] Samuel Messick. 1995. Validity of Psychological Assessment. *American Psychologist* 50, 9 (1995), 741–749.

[29] Sune Dueholm Müller, Pernille Kræmmergaard, and Lars Mathiassen. 2009. Managing Cultural variation in Software Process Improvement: A Comparison of Methods for Subculture Assessment. *IEEE Transactions on Engineering Management* 56, 4 (2009), 584–599.

[30] Wanda Orlikowski. 2002. Knowing in Practice: Enabling a Collective Capability in Distributed Organizing. *Organization Science* 13, 3 (2002), 249–273.

[31] Oleg Pankratz and Dirk Basten. 2018. Opening the black box: Managers' perceptions of IS project success mechanisms. *Information and Management* 55 (2018), 381–395.

[32] Kai Petersen and Claes Wohlin. 2009. A comparison of issues and advantages in agile and incremental development between state of the art and an industrial case. *Journal of Systems and Software* 82 (2009), 1479–1490.

[33] Kai Petersen and Claes Wohlin. 2009. Context in Industrial Software Engineering Research. In *Proceedings of the Third International Symposium on Empirical Software Engineering and Measurement (ESEM 2009)*. The Institute of Electrical and Electronic Engineers, Inc., Orlando, Florida, 401–404.

[34] Winston Royce. 1970. Managing the Development of Large Software Systems. In *Proceedings, IEEE WestCon*. The Institute of Electrical and Electronic Engineers, Inc., 328–339.

[35] Sharon Ryan and Rory V. O'Connor. 2009. Development of a team measure for tacit knowledge in software development teams. *Journal of Systems and Software* 82 (2009), 229–240.

[36] Klaas-Jan Stol and Brian Fitzgerald. 2018. The ABC of Software Engineering Research. *ACM Transactions on Software Engineering and Methodology* 27, 3 (2018), 11.1–11.51.

[37] Felix Tan and M. Gordon Hunter. 2002. The Repertory Grid Technique: A Method for the Study of Cognition in Information Systems. *MIS Quarterly* 26, 1 (2002), 39–57.

[38] Dan Tofan, Matthias Galster, and Paris Avgeriou. 2011. Capturing Tacit Archirectural Knowledge Using the Repertory Grid Technique (NIER Track). In *Proceedings of the 33rd International Conference on Software Engineering (ICSE'11)*. Association for Computing Machinery, Honolulu, HI, USA, 916–919.

[39] Rodney Turner, Ann Ledwith, and John Kelly. 2010. Project management in small to medium-sized enterprises: Matching processes to the nature of the firm. *International Journal of Project Management* 28 (2010), 744–755.

[40] Robin Whittemore, Susan K. Chase, and Carol Lynn Mandle. 2001. Validity in Qualitative Research. *Qualitative Health Research* 11, 4 (2001), 522–537.

[41] S. Michelle Young, Helen M. Edwards, Sharon McDonald, and J. Barrie Thompson. 2005. Personality Characteristics in an XP Team: A Repertory Grid Study. In *Proceedings of the 2005 Workshop on Human and Social Factors of Software Engineering (HSSE '05)*. Association for Computing Machinery, 1–7. https://doi.org/10.1145/1083106.1083123