Shared Socially Shared Regulation in Scrum Teams:

an Exploratory Study that Develops and Tests Measurement Methods

D. Pruijsers-Verduijn | S1329847

Deventer, July 2018

Examination committee Dr. M.D. Endedijk
M. Sc. M. Wijga

Project	Master Thesis
Master	Educational Science and Technology
Track	Human Resource Development
Faculty	Faculty of Behavioural, Management and Social Sciences (BMS)

UNIVERSITY OF TWENTE.

Acknowledgements

I proudly present you my master thesis. This thesis is the result of a search on how to identify the use of shared Socially Shared Regulation in Scrum teams. This study is conducted for the University of Twente, in the context of the study Educational Science and Technology.

This thesis is the result of intense work. It has made great demands to my research skills and it also took place amidst important events of my life. I got married, I gave birth to our first baby, and my father remarried. Although these events are very special, the combination entailed that it was also a hard period.

Now the result is here, I can say it brought me a better understanding on my personal abilities. Now, I use this opportunity to express my gratitude to the people who contributed to the establishment of this thesis.

I would like to thank my supervisors, PhD Marijn Wijga and dr. Maaike Endedijk, for their guidance and support during my graduation period. I would like to thank you both for your understanding attitude. Additionally, I would like to thank Marijn Wijga for making her data available for this study. I thank PhD Rike Bron for her help, on the coding schemes, during the summer holidays.

My husband, Wim Pruijsers, deserves a note of thanks: your patience, encouragement, and support at times of frustration, kept me motivated.

And, thank you Suze Pruijsers, for being such a fantastic daughter.

I hope you will enjoy reading this thesis.

Daniella Pruijsers-Verduijn,
Deventer, May 2018

Abstract

To determine if collaboration of self-directed teams is effective, measurement methods (i.e., strategies, or tools) are crucial. Although many studies emphasized the importance of (high extent) sharing of Socially Shared Regulation (SSR) for effective SSR, no studies attempted to provide an measurement method that determines the extent to which SSR is shared among team members. In our attempt to provide such an measurement method, the present study explored two measurement methods. The first method has a *qualitative approach (*focusing on *'transactive communication'),* and the second has a *quantitative* approach (focusing on *'equal team member contributions'*). To quantify the extents of sharedness of SSR, two *operationalizations* are utilized per method. The two measurement methods are tested on a series of coded videotaped meetings of self-directed teams that are guided by the Scrum Framework; an agile framework for software development. The sample included three teams that are employed at two Dutch software development organizations. Quantitative analysis of videotaped conversations indicated that the *extents of sharedness of SSR* are most successful determined by a combined measure of both methods; including one method with a qualitative approach that operationalizes the extents of sharedness of SSR based on the *durations* and *frequencies* of 'high-level transactive communication'*;* an one method with a quantitative approach that operationalizes the extents of sharedness of SSR based on the *frequencies* of 'equal team member contributions'. Analysis of the data revealed that high extents of sharedness of SSR are related to high extents of *shared conclusion-drawing*. Variations in the extents of sharedness of SSR are not found attributable to the *type of Scrum meeting*.

Keywords: measurement methods, Scrum, Social Regulation, sharedness of Socially Shared Regulation

Table of Contents

1.Theoretical Foundation 8
1.1 Regulation 8
1.2 Social Regulation (SR) 8
1.3 Sharedness of Socially Shared Regulation 9
1.4 Identifying Shared Socially Shared Regulation in the Context of Scrum 10
1.4.1 High-level transactive communication. 10
1.4.2 Equal team member contributions. 12
1.4.3 Shared Conclusion-Drawing. 12
1.5 The Scrum Framework 14
2.Method…… 16
2.1 Research Design 16
2.2 Participants 16
2.3 Materials 18
2.3.1 Videotaped meetings. 18
2.3.2 Coded video data. 19
2.4 Procedure 19
2.4.1 Construct validity process. 19
2.4.2 Measurement method development. 19
2.4.3 Coding scheme development for ‘shared conclusion-drawing’. 20
2.5 Data-Analysis 21
2.5.1 Data-analysis measurement method ‘high-level transactive communication’. 21
2.5.2 Data-analysis measurement method ‘equal team member contributions’. 23
2.5.3 Data-analysis coding scheme ‘shared conclusion-drawing’. 24
3.Results…….. 28
3.1 Descriptive statistics 28
3.2 Correlations 29
3.3 Additional Analyses 32
4.Discussion…… 38
4.1 Discussion of the Results 38
4.2 Theoretical Implications and Directions for Future Research 47

4.3 Limitations and Directions for Future Research 48
4.4 Practical Implications 50
5.Conclusion…… 51
References……. 52
Appendix A Coding Scheme ‘High-level Transactive Communication’ 56
Appendix B Raw Extents of Sharedness of SSR 58

Since the beginning of the 1990s, researchers paid increased attention to workplace learning. This expanded interest stems from rapid changes in society and working life that have taken place during the past few decades (Tynjälä, 2008). Significant changes are, for instance, globalization, innovations in the information- and communication technology (ICT); as well as new insights into workplace learning. These and similar recent changes make it necessary for organizations to adapt to these new circumstances. In response, many organizations implemented self-directed work teams (i.e., self-managing teams), in which the team is collectively responsible for goal attainment (Cohen & Bailey, 1997).

Software development organizations (SDO) similarly faced the challenge of adapting to rapid changes and complex environments (Moe, Dingsøyr, & Dybå, 2010; Rising & Janoff, 2000). As a result, during the last years, many SDO's moved away from traditional, rigid (i.e., hierarchical) organizational structures, towards more flexible (i.e., agile) ones for software development (Moe et al., 2010). Within these flexible organizational structures, collaboration success depends on the contributions of all individual group members. That is, there is both an individual, an a team accountability and responsibility (Mattessich & Monsey, 1994). This leads teams to self-organize the collaboration process.

From research, in educational contexts, we know that teams perform more effectively when joint activities are coordinated using Social Regulation (SR) (e.g., Hadwin & Oshige, 2011). SR can be understood as an overall term for regulatory processes by which collaborative work is structured (Rogat & Linnenbrink-Garcia, 2011). Typically, SR is described through iterative cycles of *planning*, *monitoring*, and *evaluating* (Zimmerman, 1989). 'Planning' involves "selecting appropriate strategies and allocating resources accordingly to organize and prepare for an upcoming task" (DiDinato, 2013, p. 27). 'Monitoring' involves "strategies individuals employ as they compare their performance with their standards or learning goals" (DiDinato, 2013, p. 27). 'Evaluation' concerns "strategies individuals use to assess learning processes and outcomes and can lead to decisions to continue, modify, or cease their actions" (DiDinato, 2013, p. 28).

In the present study, team affordances are coordinated by the *Scrum framework*. Perhaps the most interesting aspect about this framework, as a context for this study, is that it broadly follows the same iterative processes as SR. Interestingly, however, several studies showed that the effectiveness of SR of teams that work by this framework is challenged by ineffective interaction (e.g., Cho, 2008). Hence, the mere use of SR does not necessarily result in effective SR.

In literature, so called, effective SR is associated with the use of Socially Shared Regulation (SSR) (e.g., Rogat & Linnenbrink-Garcia, 2011; Volet, Vauras, & Salonen, 2009). SSR involves "constant monitoring and regulation of joint activity, which cannot be reduced to mere individual activity" (Vauras,

Iiskala, Kajamies, Kinnunen, & Lehtinen, 2003, p. 35). In literature, the importance of high extent sharing of SSR is an aspect of collaboration that is stressed frequently. Examples of aspects of SSR that indicate a certain degree of sharedness concern ‘high levels of engagement’ (Berkowitz, 1983), ‘goal-focussed talk’ (Volet et al., 2009), ‘equal team member contributions ‘ (Hoegl & Gemuenden, 2001), and ‘shared conclusion-drawing’ (Muller, 2017). We expect that mapping shared aspects of SSR is likely to enhance our understanding of interaction-related challenges in teams. In this respect, it would be useful to develop measurement methods (i.e., strategies, or tools) that determine the extent to which SSR is shared among team members. So far, however, no empirical measurement method is available. In other words, there is need for a measurement method that helps to determine the extent to which SSR is shared among team members.

The present study fills this gap in literature by addressing the following main research question: *How can the extent of sharedness of SSR in Scrum teams be determined?* As an answer to this question, we will develop two measurement methods, and assess their appropriateness to determine the extents to which SSR is shared among team members. The first method will have a qualitative approach, that examines the extents of sharedness of SSR based on the extent to which team members use *high-level transactive communication*. The second will have a quantitative approach, that examines the extents of sharedness of SSR based on the extent of *equal team member contributions*. To quantify the extents of sharedness of SSR two operationalizations will be developed per method. The operationalizations for the qualitative method focus on content-related aspects of collaboration, using coded Scrum meetings. The operationalizations for the quantitative method consider mere statistical calculations. Ultimately, we will explore how the methods are related to each other and which method, or combination (qualitative, quantitative), is most successful to represent the extents of sharedness of SSR. Research on SR can build further on the results to expand knowledge on determining the extent to which SSR is shared among team members. For practice, the results may demonstrate the need to undertake training investments that promote shared regulation among team members.

1. Theoretical Foundation

1.1 Regulation

The construct *regulation* (originally referred to as *metacognition*) is originally introduced by (Flavell, 1979). It initially referred to "cognition about cognition", which means consciously 'thinking about thinking'. Similar to "cognition about cognition", as referred by Flavell, individuals also require cognitive skills to manage their learning processes for effective knowledge acquisition. Those skills include planning, monitoring and evaluating, to which we can refer as regulation processes (Flavell, 1979). Regulation is initially aimed at facilitating a persons' knowledge acquisition (Flavell, 1979). However, more recently, the scientific focus shifted from the individual person's regulatory processes to teams' regulatory processes (Schoor, Narciss, & Körndle, 2015).

The term 'regulation' is used as an umbrella-concept for planning, monitoring, and evaluating processes, of individuals and teams, that are aimed at the attainment of a (collectively shared) goal (e.g., Hadwin & Oshige, 2011; Meijer, Veenman, & van Hout-Wolters, 2006; Schoor et al., 2015; Zimmerman, 1989). Regulation processes can be directed towards cognitive, behavioural, and motivational activities (Volet et al., 2009). In literature, three types of regulation are distinguished. Firstly, *self-regulation* which involves "the use of specified strategies to achieve academic goals on the basis of self-efficacy perceptions" (Zimmerman, 1989, p. 329). Secondly, *co-regulation* which is used to denote a collaboration of two unequal persons, whereby a more capable person supports someone in undertaking regulatory tasks (Hadwin & Oshige, 2011). Thirdly, *social regulation* which refers to collectively shared regulatory processes of teams (e.g., Schoor et al., 2015). These three types of regulation are correspondingly aimed at (shared) goal attainment (Hadwin & Oshige, 2011; Schoor et al., 2015).

The present study focussed on the investigation of regulation processes in team contexts. Hence, the study's focus is on Social Regulation (SR).

1.2 Social Regulation (SR)

The purpose of SR is to structure team affordances toward the completion of shared tasks (Schoor et al., 2015). During SR, monitoring - and control processes are iteratively discussed according to a closed feedback loop. The iterative processes are repeated until agreement is achieved (Zimmerman, 1989). These regulatory processes are expressed through individual's verbal interactions with other team members (Hadwin & Oshige, 2011; Iiskala, Vauras, Lehtinen, & Salonen, 2011). Not surprisingly, many earlier studies focused on examining collective regulatory processes though team interactions (e.g., Molenaar, 2011; Schoor et al., 2015).

Research in educational settings showed that student teams that pay attention to the discussion of these processes are more capable of effectively regulating collaborative work, than teams that not discussed those processes (Iiskala et al., 2011). This finding corresponds with previous research on collaborative learning, which demonstrated a positive relationship between high-quality interaction and effective team functioning (e.g., Molenaar, 2011). Nevertheless, although teams use SR to coordinate their team efforts, this does not necessarily assure effective SR (e.g., Vauras et al., 2003). Therefore, we advocate that the use of SR must meet certain standards to be considered *effective*: high extent sharing of SR.

1.3 Sharedness of Socially Shared Regulation

The construct SSR (i.e., SR) refers to regulatory processes of teams that are aimed at regulating joint activities in which they perform as an unit (Vauras et al., 2003). In literature, SSR is similarly defined by scholars (Schoor et al., 2015). Empirical studies unanimously agree that SSR positively affects team SR (e.g., DiDinato, 2013; Schoor et al., 2015). Indeed, many scholars proclaim that SSR is the most profound form of SR (e.g., Schoor et al., 2015; Volet et al., 2009). The use of SSR, thus seems in particularly interesting in investigating effective SR.

In SR literature, however, SR and SSR are not clearly distinguished. It, nevertheless, seems that scholars who investigated SSR in particularly focussed on the aspect of *sharedness* of regulatory processes. In doing so, verbal expressions are often used as an important starting point for measurement. This seems logical considering that regulation processes are interwoven in a persons' interaction with others (Molenaar, 2011). Examples of such shared communication-related aspects of SR concern: the extent of 'transactive communication' (Berkowitz, 1983), the use of 'goal-focussed talk' (Volet et al., 2009), 'equally shared contributions' (Hoegl & Gemuenden, 2001), 'conclusion-drawing' (van der Haar et al., 2013; Raes, Boon, Kyndt, & Dochy, 2015). To determine to what extent SSR is shared, an adequate measurement method is required. So far, however, no method is provided in literature. Consequently, in this particular study, we make an attempt to provide a valid measure that that can help to indicate the extent to which SSR is shared among team members; while focussing on two aspects ('high-level transactive communication', and 'equal team member contributions'). Additionally, the extent of sharedness of SSR is also expected to be associated with *shared conclusion-drawing*. To be able to draw a dividing line between high-, and low-extent sharedness of SSR, literature is consulted.

In short, in our attempt to constitute a valid measure for the extent to which SSR is shared, literature-based shared aspects of SSR are used to test their use as a measure.

1.4 Identifying Shared Socially Shared Regulation in the Context of Scrum

Starting from literature, three communication-related aspects are selected that, we think, are associated with the extent to which SSR is shared among team members. The first two aspects (section 1.4.1, and 1.4.2) are used as a basis for the design of the two measurement methods. The third aspect is used in our attempt to explain variations in the extents of sharedness of SSR (section 1.4.3).

1.4.1 High-level transactive communication.

The literature indicates that effective SR it is characterized by high levels of engagement (e.g., Berkowitz & Gibbs 1983; de Backer et al., 2015; Teasley, 1997). Berkowitz and Gibbs (1983), introduced the construct transactivity to denote the extent to which employees relate to, or engage in each other's regulation processes. These researchers found that high-level transactive discussions promoted effective collaboration of college-age peers. This finding correspond with the more recent work of de Backer and colleagues (2015).

In view of these studies, we argue that 'high-level transactive communication' might be related to high extent sharedness of SSR. More engaged discussions are expected to result in new, more complete, and elaborated ideas; for achieving more effective SSR. Hence, the construct of *transactivity* may be useful for studying the extents of sharedness of SSR. The literature is quite ambiguous when it comes to defining transactive communication. Berkowitz and Gibbs (1983), for instance, defined transactivity as: "reasoning that operates on the reasoning of another" (p. 402). More specifically, Teasley (1997) considered a contribution transactive if "it extends, paraphrases, refines, completes, or critiques the partner's reasoning or the speaker's own reasoning" (p. 362). In other words, in transactive discussions team members engage in, and relate to another team members' regulation activity. Although the construct is mainly used to investigate student interaction, it can also be applied at the workplace to analyse team interaction (Molenaar, 2011).

In literature several classifications of transactivity are discussed (e.g., Berkowitz & Gibbs, 1983; Molenaar, 2011; Wijga & Endedijk, 2016). Molenaars work (2011) forms an important starting point for our work. This author investigated how transactive communication is interwoven in student's interaction with others. To this end, the author distinguished between four scales of transactivity.

Verbal expressions can be either *ignored, accepted, shared,* or *co-constructed* (Molenaar, p. 115. From the context of SR, a verbal expression is 'ignored' when "the group members do not relate to nor engage in another group member's metacognitive activity (Molenaar, 2011, p. 120). SR processes are 'accepted' when "the group members engage in a metacognitive activity with a cognitive activity" (p. 120).

'Shared' regulation occurs "when a group member monitors or controls the group's learning activity and another group member relates to this activity with a metacognitive activity" (Molenaar, 2011, p. 116). Finally, 'co-construction' occurs when "group members build on each other's metacognitive activities regulating the group's cognitive activities co-constructing metacognitive activities" (Molenaar, 2011, p. 116).

The adopted coding scheme in this study (see Material section), is based on this classification of Molenaar; except that the 'sharing', and 'co-constructing' categories are merged into a new category called *engaging*. The researchers made this choice because application of the separate codes is found not appropriate for an organizational context (Wijga & Endedijk, 2016). Nevertheless, to obtain a more valid measure, we plead for a more fine-grained categorization that fits the organizational context. In this respect, the categorization of the founding study on transactive communication of Berkowitz and Gibbs (1983) would be helpful. These researchers divided the *medium* and *higher* levels of transactive communication (which roughly corresponds with the 'engaging' category) into two main categories. First, in *operational transacts* members elaboratively operate on each other's contributions by making a transformation to the other members' reasoning (e.g., through elaborative clarifying, criticising, reasoning, questioning). Second, in *representational transacts*, team members make a re-presentations of the other members' reasoning. We argue that the 'engaging' category should be divided according to this twofold categorization of Berkowitz and Gibbs (1983). The advantage their work is that they provide extensive descriptions, and examples that are applicable to SR in organizational settings.

With respect to the categorization on the levels of transactivity in the adopted coding scheme, we argue that 'engaging' utterances are associated with high extents of sharedness of SSR; because if team members build upon, and relate to each other's ideas, these are more likely to result in new information, more complete, and elaborated ideas. Correspondingly, 'engaging' communication (further referred to as 'high-level transactive communication') can be either calculated based on the *number of* (i.e., frequency) *of engaging utterances*, or *time of 'engaging' utterances*. If operationalizations are developed, based on these two indicators, SR is studied throughout the whole collaboration process; as 'engaging' communication occurs in response to a previous statement.

In conclusion, prompted by these findings, the aspect of 'high-level transactive communication' is likely to indicate the extent to which SSR is shared among team members; by focusing on content-related aspects of SSR. Possibly, we can refine the adopted coding scheme by using Berkowitz and Gibbs twofold categorization to fine-grain the 'engaging' category.

1.4.2 Equal team member contributions.

Several studies addressed the influence of team members' contributions on team performance quality (e.g., Hoegl & Gemuenden, 2001; Volet, Summers, & Thurman, 2009). Researchers on collaborative learning seem to agree that a balance in team member contributions positively affects team performance. Hoegl and Gemuenden (2001), for example, found in their study that equal contributions are associated with significantly better team performance (in terms of quality, time, and financial resources). Additionally, Volet et al. (2009) investigated university student SR patterns. The results of their study indicated that equal team member contributions are most likely to benefit the SR. A remarkable finding is that the quality of the collaboration is considerately lower when one or more member(s) remained silent.

Congruent with Hoegl and Gemuenden (2001), contributions are considered equally when "team members are contributing to the achievement of the team's goals in accordance with their specific potential" (p. 447). As aforementioned, in SSR team members take decisions as a unit (Vauras et al., 2003). In this respect, all contributions of team members can be regarded as affordances that contribute to this shared decision-making. 'Equal team member contributions' can be either operationalized based on the number of *involved team members* (Volet et al., 2009), or the *distributions of the speaking durations*.

Concluding, based on these findings, the aspect 'equal team member contributions' is likely to indicate the extent to which SSR is shared among team members. This aspect can help to determine the extent to which SSR is shared using mere a mere statistical approach.

1.4.3 Shared Conclusion-Drawing.

Previous studies suggest that effective collaborative teams draw more shared conclusions than the less effective teams (e.g., Rogat & Linnenbrink-Garcia, 2011; van der Haar et al., 2013; Muller, 2017). A study of van der Haar et al. (2013), for instance, investigated what team learning processes promoted effective team collaboration in emergency management command-and-control teams. The authors found a pattern indicating that higher performing teams use more (explicit) conclusions in their discussions than lower performing teams. Considered from a SSR context, a conclusion as a verbalized *outcome* (of a regulation activity) can allow teams to move to the next phase in the cyclical process of planning, monitoring, and evaluation. Since teams need to plan follow-up actions (or decisions) to come to concrete actions (van der Haar, et al., 2013).

But when exactly do we speak of a conclusion in the context of SR? Muller (2017) defined a conclusion as a "shared concluding utterance" (p. 20). Raes et al. (2015) added that conclusions are aimed toward an agreement, or decision that "is made concerning the topic of the (conflictual) construction" (p.

483). When considering these definitions, it seems that multiple team members come to a concluding statement about a previously discussed topic. Considered from the SR context, and in line with the above-mentioned studies, we define a conclusion as: a shared agreement, repetition, or summary on the raised regulation activity (i.e., regulation topic), or discussion.

Several classifications are available in literature for the 'types of conclusion'. The work of Raes et al. (2015) seems especially appropriate as it determines whether conclusions are mutually shared. Their classification includes three categories: *implicit*, *explicit,* and *missing conclusions* (or *wrap-ups*). An 'explicit conclusion' is expressed "when one team member expresses the conclusion and one or more other team members confirm with an affirmation" (Raes et al., 2005, p. 483). An explicit conclusion, thus involves a verbalized conclusion together with an affirmation, that indicates agreement. An 'implicit conclusion' is expressed when "one team member expresses the wrap-up and the rest does not explicitly react to counter it" (Raes et al., 2005, p. 483). The difference between the conclusions is thus determined by the presence of a confirmation. Now, conclusions can also be 'missing'. This means that no agreement is achieved, or decision is taken concerning the topic (i.e., SR activity) (Raes et al., 2015). For example, when the discussion is *interrupted,* or a conclusion is *simply missing*. Raes and colleagues also mention that conclusions can be *postponed*. For instance, when the team lacks knowledge, when essential information is missing, or when there is no immediate need to make a decision (Raes et al., 2005). This is, however, not relevant for our study.

Supported by these studies, it seems that 'shared conclusion-drawing' is an appropriate characteristic of sharing SSR. Especially 'explicit conclusions' seem to contain the high extents of sharedness; because agreement with a particular conclusion is explicitly expressed. Additionally, it also indicates whether or not the conclusion is clearly perceived. In other words, explicit conclusions help team members to gain insights into each other's thinking, which in turn enables them to more effectively plan, monitor, and evaluate their shared tasks. 'Implicit conclusions', however, may entail the risk of false interpretation; agreement is implicitly assumed. Another potential risk of implicit conclusions is that other members fear to disagree due to, for instance, a negative team climate (Raes et al., 2015). 'Missing conclusions' seem to not be associated with high extent sharing of SSR; because decisions on follow-up actions remain absent (e.g., van der Haar et al., 2013). This leads us to offer hypothesis 1.

Hypothesis 1: high extents of sharedness of SSR are positively related to high extents of shared conclusion-drawing (from low to high: 'missing conclusions', 'implicit conclusions', 'explicit conclusions').

1.5 The Scrum Framework

In this study, team affordances are coordinated by the Scrum framework. Scrum is an Agile software development process (i.e., framework) for small teams, whereby a product is developed through iterative cycles (i.e., "Sprints") of planning, monitoring, and evaluation (Rising & Janoff, 2000); a sprint (i.e., iteration) usually takes 2-4 weeks (Hossain, Babar, & Paik, 2009). Scrum and other similar Agile frameworks help organizations to quickly adapt to rapidly-changing requirements (Moe et al., 2010). In Scrum, the team is considered self-managing (Dybå et al., 2014); working towards the completion of a shared (sub)task (Moe et al., 2010).

During a Sprint cycle, several types of meetings are held. To start with, each development cycle (i.e., sprint) starts with a *sprint planning meeting*. In this start-up meeting, the team is informed about the tasks that need to be performed. Subsequently, the team determines which subtasks are performed by whom, within a certain period (Dybå, et al., 2014). Second, throughout the development process several *daily Scrum*s are organized. These are short daily-meetings in which three questions are addressed: 1. What has been done since last Scrum meeting? 2. What needs to be done before the next Scrum meeting takes place? 3. What are the hurdles? (Moe & Dingsøyr, 2008). Third, *sprint review meetings* take place to review the with respect to the software developed, technological developments, and current market conditions (Hossain et al., 2009). Fourth, at the end of the development process, a *sprint retrospective meeting* (i.e., retrospective) takes place (Moe & Dingsøyr, 2008). This final meeting enables teams to continuously improve team performance. The *retrospective* is facilitated by three main questions for discussion (Moe et al., 2010): 1. What went well during a sprint? 2. What went wrong during a sprint? 3. What can be done to improve team performance? As the types of meetings have different purpose, it is likely to assume that this might explain variations in the extents of sharedness of SSR. In line with this assumption, hypothesis 2 is stated.

No research has been done on the influence of the influence of the 'type of Scrum meeting' on the extent of sharedness of SSR. Despite important differences between the two 'types of Scrum meetings' (different meeting purpose, different duration, different number of team members involved in the meeting), this inquiry is approached with an open mindset; as it is a first attempt to explain variation by the 'type of Scrum meeting.

Hypothesis 2: differences in the extents of sharedness of SSR are attributable to the 'type of Scrum meeting' ('daily Scrums', 'retrospectives').

The present study

Many studies on SSR focused on the importance of high extent sharing of SSR for effective collaboration. Surprisingly, so far, no attempts are made to develop an empirical measurement method (i.e., analysis tool) that helps to measure the extent to which SSR is shared among team members. This gap in research motivated us to explore methods that can help to determine the extent to which SSR is shared. Additionally, SSR is mostly studied in educational contexts (e.g., Molenaar, 2011), but also plays an important role in workplace settings (Schoor, et al., 2015). Therefore, context-specific research is needed to identify context-specific SR, rather than simply assuming context homogeneity. Uniquely, this study presents two methods with a different approach; one method with a 'qualitative', and one method with a 'quantitative' approach. The method with the qualitative approach determines the extents of sharedness of SSR, based on the extent to which team members use 'high-level transactive communication'. The method with the quantitative approach determines the extents of sharedness of SSR, based on the extent to which team members provide 'equal team member contributions'. To quantify the extents of sharedness of SSR two operationalizations are developed per measurement method. The operationalizations for the qualitative method focus on content-related aspects of collaboration, using coded Scrum meetings. The operationalizations for the quantitative method consider a mere quantitative approach, using statistical calculations. To determine which method, or combination (qualitative, quantitative) is most successful in determining the extents of sharedness of SSR, the outcomes of the operationalizations will be presented, compared, interpreted, and explained; while respecting the quality criteria (section 2.4.2). Correspondingly, the present study seeks to answer the following main question.

RQ: How can the extent of sharedness of SSR in Scrum teams be determined?

Subsequently, the external validity of the measurement methods is examined; to explain variations in the extents of sharedness of SSR. For that purpose, the extents of sharedness of SSR are compared to, and explained by the extent to which conclusions are shared. As aforementioned, it is expected that high extents of sharedness of SSR are related to high extents of shared conclusion-drawing. This examination is guided by sub-question a.

SQ a: Can high extents of sharedness of SSR be attributed to high extents of shared conclusion-drawing (from low to high: 'missing conclusion', 'implicit conclusion', and 'explicit conclusion')?

In addition, variations in the extents of sharedness of SSR are by explained by, and compared to the ‘type of Scrum meeting’. This inquiry is addressed by sub-question b.

SQ b: Can variations in the extents of sharedness of SSR be attributed to the ‘type of Scrum meeting’ (‘daily Scrum’, and ‘retrospective’)?

The ‘sprint planning’, and ‘sprint review meetings’ are excluded from analysis. For the ‘sprint planning meetings’, insufficient data are available to answer this research question. The ‘sprint reviews’ are excluded because external parties are involved in the meeting that may distort the results.

The findings of this study contribute to both empirical research, and organizational practices. This study extends empirical research on communication-related aspects that aim to determine the extent of sharedness of SSR. This study provides room for other authors to confirm, improve, expand, and further tighten the methods (e.g., operationalizations). This study also contributes to practice, in that the methods can help both software development organizations, and their Scrum teams to gain insights in the sharedness of their SSR process. Accordingly, the results may demonstrate the need for organizations to invest time, and resources necessary to stimulate shared regulation among team members (e.g., through on-the-job training); to contribute to more effective SSR.

2. Method

2.1 Research Design

This study has characteristics of both an *exploratory*, and *observational study design*. The exploratory study design is utilized in the exploration of what aspects of SSR constitute a valid measure for the sharedness of SSR. Additionally, the observational study design is employed to observe the characteristics of shared aspects of SSR; without intervening at the workplace. Ultimately, the study generated quantitative data.

2.2 Participants

The data-collection occurred at two Dutch software development organizations. For privacy reasons, we call these organizations, ‘organization 1’, and ‘organization 2’. The sample included 16 software engineers. In Table 1, the demographics of the participants are presented.

Table 1

Demographics of Sample per Team

Demographic Construct	Measure	Team A	Team B	Team C
Team Size	*f*	5	6	5
Educational level	*f*	5	6	5
Bachelor	*f*	3	4	3
Master	*f*	2	2	2
Age	*M*	45	40	36
	SD	8.05	5.12	1.76
Gender	*f*	5	6	5
Female	*f*	1	0	0
Male	*f*	4	6	5
Function	*f*	6	7	8
Developer	*f*	3[a]	4[a]	3[a]
Functional Designer	*f*	0	1	0
Product Owner	*f*	1	1	0
Scrum Master	*f*	1	0	1
Solution Architect	*f*	1	1	0
Tester	*f*	0	0	4[a]
Nationality	*f*	5	6	5
Dutch	*f*	4	6	5
Non-Dutch	*f*	1	0	0
Period current team formation	*f*	1[b]	2[b]	1[b]

Note. These data are obtained by Marijn Wijga, through a questionnaire on team member characteristics.
[a]Four participants represented double roles. [b]In months.

The participants are spread out across three teams, that are called “team A”, “team B”, and “team C”. The average age of the participants is 40.25 years (*SD* = 6.82); ranging from 33 to 60 years. An one-way analysis of variance showed that the differences in age are statistically non-significant, $F(2, 13) = 2.56$, $p < .10$. The respondents are predominantly male (94%). Of the total sample, the majority was in possession of a bachelor's degree (62.5%). The members fulfilled the following team roles: ‘developer’, ‘functional designer’, ‘product owner’, ‘Scrum master’, ‘solution architect’, or ‘tester’. With respect to the team roles, three members represented double roles, whereas one member represented three roles. A striking point is that all three Scrum masters are charged with double roles. Which is uncommon because

the Scrum master is supposed to represent a single, and independent role as coach (Moe & Dingsøyr, 2008). Lastly, looking at the origin of the respondents, the majority is Dutch (94%); only one member of team A came of outside the Netherlands. All participants voluntarily participated, and received no monetary compensation for participation.

In conclusion, the differences in team member characteristics are rather small. Therefore, it is not considered necessary to investigate variations in the extents of sharedness of SSR across the three teams.

2.3 Materials

2.3.1 Videotaped meetings.

The data are collected by Marijn Wijga; a PhD-level researcher. This researcher observed and recorded series of sprint planning -, daily Scrum -, sprint review -, and sprint retrospective meetings. She made her data available for our study before the study start date. These data are part of a larger research project conducted at the University of Twente. This part of the study focused on SR. The data consist of 25 audiovisual recordings of Scrum meetings; resulting in over 6 hours of videotaped data. It contained 1.674 utterances, spread over 213 episodes. In Table 2, the number of recordings per team, and 'type of Scrum meeting' are displayed. The average duration of a daily Scrums is 8 minutes, and 37 minutes on average for the retrospectives.

Table 2

Number of Regulation Episodes and Audiovisual Recordings per Team and Type of Scrum Meeting

	Team A		Team B		Team C	
Type of Scrum meeting	Episodes	Recordings	Episodes	Recordings	Episodes	Recordings
Daily Scrum	49	9	44	5	30	5
Retrospective	39	2	25	1	26	3
Total	88	11	69	6	56	8

Note. We tried to equally distribute the number of episodes, and videos across the various teams and meetings. Due to a lack of available video observations, we are not entirely able to analyse equal amounts of episodes and meetings.

2.3.2 Coded video data.

To test our measurement methods, we used coded video data. Before the study start data, we received the coded data in Observer® XT 13. The data are coded using the coding scheme of Muller (2017, see Appendix A, Table 9). Originally, this coding scheme is developed by Wijga and Endedijk (2016). Muller modified the coding scheme in such a way that it allowed for determining SSR; based on shared conclusion-drawing. We further extended the coding (see Data-Analysis section). To enhance consistency in the application and interpretation of the current coding, we recoded four videos. Observer agreement is calculated with Cohen's Kappa. Assuming Landis' and Koch's (1977) classification for the reliability of Kappa values, a satisfying score of $K = .80$ ($p < .01$) is reached; indicating a strong agreement. The fidelity of Muller's coding scheme is 0.914.

2.4 Procedure

2.4.1 Construct validity process.

Construct validity is a validity measure that is defined as "the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured" (Carmines & Zeller, 1979, p. 23). To test the construct validity of our two methods, three steps are established. First, the theoretical relationship between the aspect of sharing SSR (with operationalizations) and the measurement methods is explained (see also Theoretical section). Second, the statistical correlation between the two methods is investigated. As a final step, quantitative findings are presented, compared, interpreted, explained, and possible scientific evidence on the relationship between the shared aspects of SSR is provided (Carmines & Zeller, 1979).

2.4.2 Measurement method development.

For this study we developed two methods that determine the extent of sharedness of SSR. In developing these methods, we faced the challenge of how to provide valid, and reliable measures. For that purpose, *quality criteria* are defined. First, the methods require theoretical substantiation. Second, the methods should be appropriate to perform measures on coded video data. Third, the methods are deemed to be sufficiently discriminative. Fourth, the methods should be cost-effective meaning the most representative results against the lowest costs (e.g., time, financial resources, and effort). The decisive factor for the most appropriate measurement method, is the extent to which the operationalizations correlate on the extents of sharedness of SSR. A method is considered appropriate if it (moderately) correlates with at least one other operationalization. This can either concern an operationalization of the qualitative, or the

quantitative method. To systematically review the suitability of the methods as measures for the extents of sharedness of SSR, these quality criteria are regarded as a minimum standard.

2.4.3 Coding scheme development for 'shared conclusion-drawing'.

To examine the external validity of the results, the extent of 'shared conclusion-drawing' (as a characteristic of team collaboration) is used to explain variations in the extents of sharedness of SSR. Correspondingly, we developed a coding scheme, to classify the episodes into three distinct types of conclusions. The coding process included three general steps.

First, a list with initial coding categories are drafted.

Second, the coding categories are tested and revised. For that purpose, we applied the coding categories to a first round of coding, using sample scripts. This first round of coding, quickly demonstrated the need to improve the initial categories to fit our study purpose and research questions. This resulted in a new set of coding categories, descriptions, and examples that are assessed by the study supervisor. Then, multiple feedback sessions took place for refinement of the codes. After that, we and the study supervisor independently double-coded a set of data. The size of this sample is calculated using Kappa analysis, general guidelines: indicating that 15 – 20 % of the total sample should be double-coded for Kappa coefficient analysis (Klonek, Quera, Burba, & Kauffeld, 2016). The sample size met this standard. Then, the coded data-sets are compared, and thoroughly discussed. At that point, the coding categories underwent a final refinement, and resulted in a satisfactory set of coding categories, with a Cohen's Kappa value of .82 (Cohen, 1960).

Third, the agreed-upon codes are assigned to the video data; using Observer® XT 13. During coding, memos are written in a log book; describing ambiguous, unclear, and doubtful codes. The memos are jointly discussed with the study supervisor, until agreement is established.

2.5 Data-Analysis

In this section, the two measurement methods are presented. First, the qualitative method quantifies the extents of sharedness of SSR based on the aspect 'high-level transactive communication' (section 2.5.1). Second, the quantitative method examines the extents of sharedness of SSR based on 'equal team member contributions' (section 2.5.2). Additionally, we attempt to explain variations in the extents of sharedness of SSR based on our coding scheme 'shared conclusion-drawing' (section 2.5.3). But before going into detail about the methods, we make four general comments.

First, we underline that, it is not our intent here to make generalizations about extent to which SSR is shared in Scrum teams. Rather, we use this opportunity to compare, interpret, and explain measurement methods that aim to determine the extents to which SSR is shared among team members.

Second, we do not have professional knowledge regarding software engineering. Consequently, we are not able to assess the discussions on substantive correctness. Though the involved software engineers are considered experts within their domain; it is thus surmised that the substantive information is correct.

Third, we study SR processes at the metalevel. In doing so, we focus on verbal statements of teams, the *unit of analysis* is on three groups of Scrum teams.

Fourth, we define an *utterance* as one or more contiguous statement(s) of an individual, referring to a single topic, or discussion. We stress out that the adopted coding scheme did not include a definition of an utterance (see Appendix A, Table 9). Aware of this fact, an utterance is defined in such a way that it is not compromised by the current coding. Nevertheless, we use Molenaars' definition of an episode: "a sequence of utterances about the same topic or discussion" (Molenaar, 2011). Thus, a new episode starts, with the first sentence of a topic. The start of a new episode is marked by the "initiating" code. No separate code is used to indicate the end of an episode; as the end is marked by the last utterance on the same topic.

2.5.1 Data-analysis measurement method 'high-level transactive communication'.

This first method, has a qualitative approach to analyze SSR. To generate quantitative data, the aspects 'high-level transactive communication' is translated into quantifiable extents of sharedness of SSR. For that purpose, two operationalizations are developed. Consistent with the theoretical foundation, 'high-level transactive communication' is supposed to be linked to high extent sharedness of SSR. Correspondingly, it is our assumption that "engaging" utterances are linked 'high-level transactive communication'; because employees engaging in, and relate to each other's SR activities.

Coding scheme. The coding scheme of Muller (2017) is used to classify the extents of 'transactive communication'. For the coding scheme, with descriptions, and representative examples see Appendix A.

Code explanation. The adopted coding scheme, categorized transactivity into four main categories from low- to high-transactive: “initiating”, “ignoring”, “accepting”, and “engaging”. As we posited in the theoretical discussion, this current categorization is rather coarse-grained for the “engaging” category; that covers a broad spectrum of both high- and low extent of sharedness of SSR. This may result in a less valid measure. This led us to plead for a more advanced categorization. In doing so, we sought to apply Berkowitz and Gibbs (1983) twofold categorization to split the “engaging” category into two: ‘operational transacts’, and ‘representational transacts’. In our attempt to make this distinction clear, we did, however, not succeed. As a consequence, we apply the original coding categories.

Codes assignment. The transactivity codes are (for the most part) assigned by other researchers. Nevertheless, we found that codes in the dataset were missing, and were not always consistent with our believes. To this end, changes were made to the current coding; in consultation with the study supervisor.

Operationalizations. The data-analysis focused on the episode level-analysis; the episode boundaries are marked by the “initiating” code (Appendix A, Table 9). In contrast to the operationalizations of the quantitative method (presented in section 2.5.2), the data of this method is analyzed using a coding scheme. The transactivity codes, are translated into *relative quantifications;* indicating the extents of sharedness of SSR per episode. In line with the theoretical framework, the extent of ‘high-level transactive communication’ is operationalized based on the *frequency*, and the *duration of engaging utterances.* The ratios of both operationalizations ranged from zero to one. Low ratios represented low extents of ‘high-level of transactive communication’, evolving low extents of sharedness of SSR; and vice versa.

Operationalization A: Relative duration HLTC (High-Level Transactive Communication). In this operationalization, the absolute total duration (in seconds) of the “engaging” utterances, within a particular episode, is divided by the total SR episode duration (in seconds); resulting in a *relative quantification of engaged time*.

Operationalization B: Relative frequency HLTC (High-Level Transactive Communication). In this operationalization, the absolute number of “engaging” utterances is summed per episode, and then divided by the total number of SR utterances within an episode; resulting in *relative quantification of engaged turns*.

Measurement method validation. To determine whether, and to what extent, there exists a possible relationship between the operationalizations of this qualitative method, and those of the quantitative method (section 2.5.2), a correlational table is constructed. Additionally, we examined to what extent variations in the extents of sharedness of SSR are related to the extents of ‘shared conclusion-drawing’ (see section 2.5.3), and to the ‘type of Scrum meeting’.

2.5.2 Data-analysis measurement method 'equal team member contributions'.

This second method, focusses on quantitatively analyzing SSR. To generate quantitative data, the investigated aspect of sharing SSR ('equal team member contribution') is translated into quantifiable extents of sharedness of SSR. For that purpose, two operationalizations are developed. Our study of the literature showed that a balance of team member contributions seems to be associated with high extent of sharedness of SSR. This stems from the idea that equal contributions stimulate team members to take decisions as an unit. Therefore, to be able to speak about high extent sharing of SSR, it is assumed that all team members should be engaged in the regulation activity (i.e., topic) equally.

Operationalizations. The aspect 'equal team member contributions' will be translated into quantifiable extents of sharedness of SSR, using mere *statistical calculations*. Instead, of using a coding scheme. In accordance with the theoretical framework, this aspect will be quantified based on the *number of involved team members*, and *the distribution of the speaking rates.* The operationalizations are utilized per episode; the episode boundaries are marked by the "initiating" code (see Appendix A, Table 9). The scores of both operationalizations ranged from zero to one. Low ratios represented low extents of 'equal team member contributions', evolving low extents of sharedness of SSR; and vice versa.

Operationalization C: Relative duration CTM (Contributing Team Members). In this operationalization, the durations of the SR time involved (in seconds) are summed per participant, per episode. Then, the sum of the shortest time involved per episode is divided by the sum of the longest time involved per episode. The outcomes indicate the extent of 'equal team member contributions' in terms of *relative quantifications of time involved*. If only one person is involved in an episode, the episode received a zero score.

Operationalization D: Relative number CTM (Contributing Team Members). In this operationalization, the absolute number of involved team members (in a particular episode), is divided by the total number of members present at the meeting. The outcomes indicate 'equal team member contributions' in terms of *relative quantifications of team members involved*. A member is considered involved if a (verbal) contribution of at least one second is coded.

Measurement method validation. A correlational table is constructed to determine whether, and to what extent, a relationship exists between the operationalizations of this quantitative method, and the operationalizations of the qualitative method (section 2.5.1). We also performed statistical analysis to determine to what extent variations in the extents of sharedness of SSR are linked to the extent of 'shared conclusion-drawing' (see section 2.5.3), and to 'the type of Scrum meeting'.

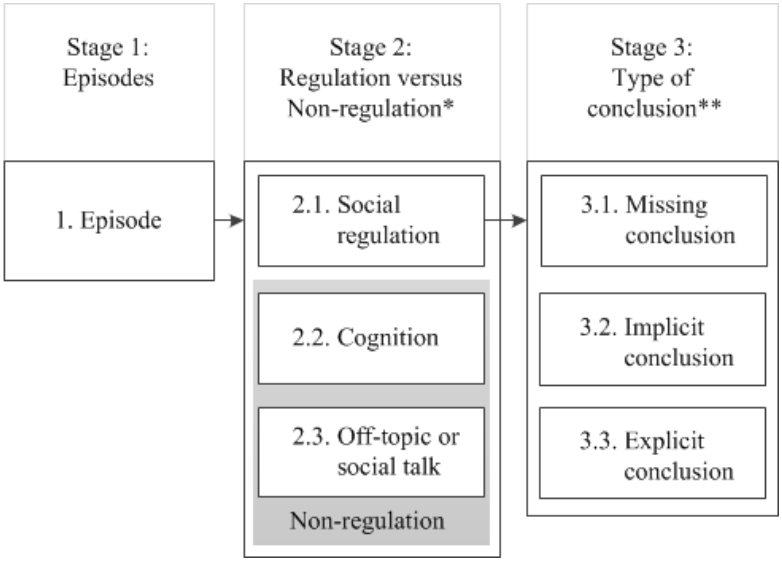
2.5.3 Data-analysis coding scheme 'shared conclusion-drawing'.

To explain variations in the extents of sharedness of SSR, a coding scheme is developed; that classified the episodes into extents of 'shared conclusion-drawing'. In line with the theoretical framework, we assume that high extent conclusion-drawing is associated with high extents of sharedness of SSR; because a conclusion, as an outcome of a SR, is expected to help teams to move to the next phase in the cyclical process of planning, monitoring, and evaluation.

Coding process. This coding scheme classified the episodes into three distinct types of conclusions: 'missing conclusion', 'implicit conclusion', and 'explicit conclusion'. The coding process is depicted in Figure 1. As the Figure illustrates, Stage 1, and 2 correspond with the coding scheme of Muller (2017).

Figure 1

Schematic Representation of the Coding Scheme



Note. The extents of sharedness of SSR are coded in three Stages. In Stage 1, the episode boundaries are marked. In Stage 2, SR utterances are distinguished from non-SR utterances. In Stage 3, each utterance received a conclusion code.
*Coded at the utterance level. **Coded at the episode level.

Code explanation. We understand a conclusion in general as a shared agreement, repetition, or summary on a raised regulation activity (i.e., topic), or discussion. A shared agreement must be understood as a *command* for a team member to take an action or to perform a task, or an *affirmation* of a team member to perform an *action* which is related to the regulation activity of conversation, or a *decision* or *compromise* that closes the topic or problem raised. When the conclusion concerns an *action*, this should be translated into a workable plan that allows team members to accomplish tasks. As aforementioned, Raes et al.'s (2015) conceptualization is used to categorize the extents of 'shared conclusion-drawing'.

If a conclusion remains absent the 'missing conclusion' code is assigned (see Table 3). This code is assigned when an episode is *interrupted*, or when a conclusion is *simply missing*. An episode is considered interrupted when it is intervened by another subject (see example 1.2 in Table 3). If such an interrupted episode is resumed in a later episode, the initial 'missing conclusion' code remained valid. An utterance is coded as 'implicit conclusion' (see Table 3), if a confirming response (see example 2.1 in Table 3) is not forthcoming after the conclusion is announced. An utterance is coded as '*explicit conclusion'* (see Table 3) if the conclusion is confirmed by one, or more team members (see examples 3.1; 3.2; 3.3; 3.4; 3.5 in Table 3). It is important that this confirmation is related to the conclusion expressed (see example 1.1 in Table 3).

The code assignment for 'retrospectives' deserves an important comment. As explained in the Theoretical section, these meetings focus on questions on what went well/wrong during a sprint, and what can be improved in future sprints (Moe et al., 2010). This implies that sharing certain information can also be sufficient to be considered as a conclusion. However, for this information-sharing to be considered as a conclusion, we argue that conclusion of these episodes should include: *descriptions*, *examples*, *explanations*, or *reasons* on what went well, or wrong (see examples 1.3; 1.4; 3.6 in Table 3).

Code assignment. The conclusion codes are coded at the episode level of analysis. The episode boundaries are indicated by the "initiating" codes. Each episode received one conclusion code; the 'implicit, and 'explicit' conclusion codes are assigned to the utterance that included the conclusion. The 'missing conclusion' code, is assigned to the final utterance on the same topic. Conclusions are not necessarily announced at the end of an episode. Lastly, we considered the fact there is an overlap between the different types of conclusions. An 'explicit conclusion' also consists of an 'implicit conclusion'. If multiple codes are applicable, the code with the highest extent of sharedness of SSR is assigned. Humor and jokes are not regarded as a conclusion; because these are expected to not contribute to tasks accomplishment.

Result analysis. To test hypothesis 1, we carried out a series of statistical tests (independent sample t-tests, and a Mann–Whitney U-test). With the aim to determine whether variations in the extents of sharedness of SSR, can be attributed to the extent of 'shared conclusion-drawing' (independent variable).

Table 3

Codes Extents of Shared Conclusion-drawing with Definition and Example

Code	Definition	Example
Missing conclusion	Regulation episodes that do not end with a verbalized conclusion on the content of the topic discussed. (E.g., no conclusion present, interrupting a conversation by introducing another topic)	<u>Example 1.1: missing conclusion; non topic-related answer</u> TM a: *"Are there other things that we need to discuss now?"* TM b: "*No impediments."* (TM b attempts to make a humorous contribution referring to a joke expressed earlier in the meeting) TM a: *"Oké cool, then we will now continue our work."* (TM a states as a conclusion to continue with the work after the joke of TM b. The initial question remains unanswered) <u>Example 1.2: missing conclusion; interrupted episode</u> TM a/b/c/d: (the team members engage in a discussion on the project) TM b: *"O, my wife needs to go away this afternoon, so I will work at home from 10 a.m. till 12 p.m."* TM a/b/c/d: (do not further discuss the initial topic) (TM b **interrupts** the discussion, by making an off-topic comment) **Retrospective information sharing** <u>Example 1.3: missing conclusion</u> TM a: *"What went well this Sprint?"* TM b: *"The refinement went well."* (TM b shares what he thinks went well, however he provides no description, example, explanation, or reason for this opinion) <u>Example 1.4: missing conclusion</u> TM a: *"What went well this Sprint?"* TM b: *"Everything went good."* (TM b shares what he thinks went well, however he provides no description, example, explanation, or reason on what went well for this opinion)
Implicit conclusion	Regulation episodes that end with a verbalized conclusion on the content of the topic discussed, which are not confirmed by one or more other team members.	<u>Example 2.1: implicit conclusion</u> TM a: *"Oke', then I suggest we continue with the next subject, and discuss this at the next meeting."* TM b/c/d: remain silent (TM a takes de decision to postpone subject discussion, **without being confirmed** by another member)
Explicit conclusion	Regulation episodes that end with a verbalized conclusion on the content of the topic discussed, which are confirmed by one or more other team members	<u>Example 3.1: explicit conclusion; with affirmation to take an action</u> TM a/b/c/d: (team members conduct a discussion about whether or not to add a new story to the Sprint) TM a: *"I'll let us make a new story. Then, we can discuss it with the team and then I think the story will be added in this Sprint."* (TM a **affirms** to take an action, which **is** not **confirmed** by another member) <u>Example 3.2: explicit conclusion; with affirmation to take an action</u> TM a: *"So, I can just start."* (TM a expresses an **agreement** on the topic discussed) TM b: *"Yes."* (TM b **confirms the agreement** expressed by TM a)

Example 3.3: explicit conclusion; command to take an action
TM a: *"This has a high priority. Do you want to do that TM b?"*
TM b: *"Yes, that's fine by me."*
(TM b agrees to take the command as given by TM a)
Example 3.4 explicit conclusion; majority voting process
TM a/b/c/d: (team members vote on whether to add more tasks to the sprint.
(TM a expresses the **agreement** as the result of a **voting process**)
Example 3.5: explicit conclusion; postponing a decision
TM a: *"Let us look at this later."*
(TM a **suggests** to return to the subject later on)
TM b: *"Yes, indeed, we must first know if Jan agrees with this."*
(TM b provides a **reason** for postponing)

Retrospective information sharing

Example 3.6 explicit conclusion
TM a: *"I think it is good that we have decided to stop with new stories."*
TM a continues: *"Therefore, I could make the art stable."*
(TM a expresses what he thinks went well, he also provides a **reason** on why he thinks it was a good decision to stop with new stories)

3. Results

In total, 213 episodes, and 1.674 utterances are identified. For the raw extent of sharedness of SSR, Table B1 in Appendix B can be consulted. The results are presented per research question.

3.1 Descriptive statistics

Before going into detail about the research questions, the descriptive statistics on the extents of sharedness of SSR are presented Table 4. It shows that the operationalization 'Relative duration HLTC' (A) has the highest median (*Mdn* = .66) of sharedness of SSR, whereas operationalizations 'Relative duration CTM' (C) has the lowest median (*Mdn* = .13). The median extents of sharedness of SSR are similar for three of the four operationalization : 'Relative duration HLTC' (A), 'Relative frequency HLTC' (B), and 'Relvative number CTM' (D).

Table 4

Descriptive Statistics for the Extents of Sharedness of SSR per Operationalization

Operationalization	*N*	*M*	*Mdn*	*SD*	Skewness	Z Skewness	Kurtosis	Z Kurtosis
Rd HLTC (A)	213	.53	.66	.35	-.0491	-2.94	-1.306	-3.93*
Rf HLTC (B)	213	.50	.60	.31	-0.652	-3.90*	-0.959	-2.89
Rd CTM (C)	213	.22	.13	.24	1.297	7.766*	0.978	2.946
Rn CTM (D)	213	.63	.67	.26	0.024	0.14	-1.059	-3.19

Note. Statistical significance is tested at the $\alpha = 0.05$ level. *Z-scores that exceeded the critical value of ±3.29 for normality. Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members.

In addition, in Table 5 the medians and the standard deviations are presented per operationalization; split by the 'type of Scrum meeting'. It can be seen that, that the median extents of sharedness of SSR are not significantly different for 'daily Scrum' and 'retrospectives'. Except for the operationalization 'Relative duration CTM', that has a lower overall median of sharedness of SSR (*Mdn* = .16).

Table 5

Extents of Sharedness of SSR per Operationalization per Type of Scrum Meeting, Team and Overall

		Sharedness of Socially Shared Regulation							
		High-level transactive communication				Equal team member contributions			
		Rd HLTC (A)		Rf HLTC (B)		Rd CTM (C)		Rn CTM (D)	
Type of Scrum meeting	*N*	*Mdn*	*SD*	*Mdn*	*SD*	*Mdn*	*SD*	*Mdn*	*SD*
Sprint planning	123	.64	.33	.57	.28	.16	.30	.67	.24
Retrospective	90	.68	.39	.60	.34	.13	.27	.50	.28

Note. Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members.

3.2 Correlations

To address the general research question of this study a series of correlational tests are performed.

RQ: How can the extent of sharedness of SSR in Scrum teams be determined?

To examine if there are any correlational relationships between the two measurement method (and thus the four operationalizations), a series of Spearman rank-order correlations are calculated. Prior to analysis, the data is examined for Pearson's *r* assumptions of homoscedasticity, linearity, and normality. The data satisfied the assumption of homoscedasticity; indicated by linear regression analysis. Unfortunately, the data violated the assumptions of linearity, and normality (Table 4). To this end, we decided to report the non-parametric Spearman's Rho correlations. Beforehand, we also calculated the correlational relationships with the parametric Pearson's *r* test. The test results revealed similar correlations for the extents of sharedness of SSR. This is believed to result higher validity of the correlations; since not severely skewed data, do not disregard the use of the Pearson's *r* correlation (Chok, 2010).

The Spearman's Rho correlations are displayed in Table 6. It shows, that all correlations are significant positively correlated, which suggests a certain relationships between the operationalizations. Therefore, all correlations are discussed.

Table 6

Spearman Intercorrelations on the Operationalizations as a Function of the Extent of Sharedness of SSR

Operationalization	*1.*	*2.*	*3.*	*4.*
1. Rd HLTC (A)	-			
2. Rf HLTC (B)	.841	-		
3. Rd CTM (C)	.236	.256	-	
4. Rn CTM (D)	.546	.572	.205	-

Note. All correlations are significant at the 0.01 level (2-tailed). Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative duration Contributing Team Members.

First, the statistical relationship between operationalization 'Relative duration HLTC' (A), and 'Relative frequency HLTC' (B) is examined. The two-tailed Spearman's test of significance yielded a positive correlation $r_s(213) = .841$, $p < .001$ between the operationalizations. This suggest that the higher extents of sharedness of SSR are for the operationalization 'Relative duration HLTC' (A), the higher the extents are for the operationalization 'Relative frequency HLTC' (B). For interpreting the effect size of correlation coefficients, we use the Rule of Thumb of Hinkle, Wiersma, and Jurs (1988). Effect size values are *negligible* between .00, and .30; *low* between .30, and .50;. *moderate* between .50, and .70; *high* between .70, and .90; *very high* between .90, and 1. There is thus a 'high' positive correlation between the operationalizations.

Second, the same two-tailed Spearman's test of significance is used to determine the statistical relationship between the operationalizations 'Relative duration HLTC' (A) and 'Relative duration CTM' (C). The Spearman's Rho reported a positive correlation of $r_s(213) = .236$, $p < .001$. Using Hinkle et al.'s (1988) Rule of Thumb for effect sizes. The operationalizations show a 'negligible' correlation.

Third, the Spearman's Rho also found a positive correlation $r_s(213) = .546$, $p < .001$ between the operationalizations 'Relative duration HLTC' (B) and 'Relative number CTM' (D). The Rule of Thumb indicates a 'moderate' effect size for this correlation (Hinkle et al., 1988).

Fourth, the Spearman's Rho test between the operationalizations 'Relative frequency HLTC' (B) and 'Relative duration CTM' (C) revealed a positive correlation $r_s(213) = .256$, $p < .001$. This effect size value is, however, 'negligible' (Hinkle et al., 1988).

Fifth, the correlation measure between the operationalizations 'Relative frequency HLTC' (B) and 'Relative number CTM' (D) indicated a positive statistical association $r_s(213) = .572$, $p < .001$ between the operationalizations. The effect size value indicated a 'modest' correlation (Hinkle et al., 1988).

Sixth, the statistical correlation between the operationalizations 'Relative duration CTM' (C) and

'Relative number CTM' (D) is examined. The correlation measure yielded a positive correlation $r_s(213) = .205$, $p < .001$ for the extents of sharedness of SSR. This effect size is, however, considered 'negligible' (Hinkle et al., 1988).

To obtain an overall correlation between the four operationalizations, the *mean correlation* is calculated. This test is done, to obtain a general indicator of whether the four operationalizations (and thus the two measurement methods) combined, are sufficiently correlated to be a valid measure for the extent of sharedness of SSR. The mean correlation is $r_s = .44$, which means that the four operationalizations are 'low' positive correlated (Hinkle et al., 1988). The correlations in Table 5, show that all operationalizations yielded a 'low' correlation with the operationalizations 'Relative duration CTM' (C). To this end, we added an additional 'mean correlation' that leaves this operationalization out of calculations; resulting in a 'modest' mean correlation between the operationalizations 'Relative duration HLTC' (A), 'Relative frequency HLTC' (B), and 'Relative number CTM' (D) ($r_s = .65$).

Respecting the quality criteria presented in section 2.4.2, both the presented measurement methods seem appropriate to determine the extents of sharedness of SSR. Only the operationalization 'Relative duration CTM' (C), however seems less appropriate as a measure because it does not sufficiently correlate with at least one other operationalization. It is, therefore, expected that this operationalization of the quantitative method is less suitable as a measure for the extents of sharedness of SSR.

3.3 Additional Analyses

Testing Hypothesis 1. To test hypothesis 1, and examine the external validity of our results, SQ a. is addressed.

SQ a: Can high extents of sharedness of SSR be attributed to high extents of shared conclusion-drawing (from low to high: 'missing conclusion', 'implicit conclusion', and 'explicit conclusion')?

As an answer to this question a series of statistical analysis are performed. In the analyses, we used an $\alpha = .05$ as a threshold for statistical significance. The independent variables represented the three extents of shared conclusion-drawing ('missing conclusion', 'implicit conclusion', 'explicit conclusion'). The dependent variables are the extents of sharedness of SSR utilized by the four operationalizations. Before going into detail about the test results, Table 7 gives the descriptive statistics on the extents of sharedness of SSR per operationalization; classified by the extent of 'shared conclusion-drawing'.

Table 7

Descriptive Statistics for the Extent of Sharedness of SSR per Operationalization and Type of Conclusion

Operationalization	Type of Conclusion	*N*	*M*	*Mnd*	*SD*	Skewness	Z Skew	Kurtosis	Z Kurtosis
Rd HLTC (A)	Missing	65	.42	.54	.38	0.091	0.24	-1.774	-3.03
	Implicit	35	.59	.69	.32	-0.912	-2.29	-0.420	-0.54
	Explicit	113	.58	.70	.33	-0.684	-3.01	-0.919	-2.04
Rf HLTC (B)	Missing	65	.40	.50	.35	-0.114	-0.38	-1.726	-2.95
	Implicit	35	.54	.56	.28	-0.984	-2.47	0.033	0.04
	Explicit	113	.54	.60	.28	-0.894	-3.94*	-0.234	-0.52
Rd CTM (C)	Missing	65	.20	.13	.24	1.428	4.81*	1.430	2.44
	Implicit	35	.17	.12	.19	1.179	2.96	0.419	0.54
	Explicit	113	.24	.14	.25	1.297	5.35*	0.683	1.51
Rn CTM (D)	Missing	65	.54	.50	.29	0.295	0.99	-1.117	-1.91
	Implicit	35	.62	.60	.27	0.209	0.53	-1.371	-1.76
	Explicit	113	.69	.67	.23	0.043	0.19	-1.009	-2.24

Note. * Z-scores that exceeded the critical z-score ±3.29 for normality. Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members.

To determine significant differences in the extents of sharedness of SSR, either a Welch, or a Kruskal-Wallis H test is performed; depending on the confirmation of the assumptions. The test scores are reported in Table 8.

Table 8

Analysis of Variance for Standardized Test Scores

Operationalization	Source	*F-stat*	χ^2	*df1*	*df2*	*p*
Rd HLTC (A)	Welch	4.88	-	2	88.98	.010
Rf HLTC (B)	Kruskal-Wallis	-	4.643	2	-	.098
Rd CTM (C)	Kruskal-Wallis	-	4.056	2	-	.132
Rn CTM (D)	Welch	6.123	-	2	82.91	.003

Note. 0.05 significance level is used. Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members; F-stat = *F* statistic.

First, the descriptive data of operationalization 'Relative duration HLTC' (A) (Table 7), show that 'implicit conclusions' (M = .59) have the numerically highest mean extent of sharedness of SSR, and 'missing conclusions' the smallest mean level (M = .42). Before analyses, the assumptions of normality, and equal variances are checked. The normality assumption is tested, using skewness and kurtosis (e.g., Mardia, 1970). The z-scores for skew and kurtosis did not exceed the absolute critical value of ±3.29 for medium-sized samples ($50 < n < 300$) (Kim, 2013). Hence, the data satisfied the normality assumption. The variances are compared using Levene's test. The test result indicated significantly heterogeneous variances (F = 2, 210) = 6.12, p = .003). Since we violated the assumption of equal of variances, the Welch's ANOVA is required, with an adjusted F ratio (10.95). The Welch's ANOVA revealed a statistically significant test result $F(2, 88.98) = 4.88$, p = .01. Therefore, we reject the null hypothesis, and accept the alternative hypothesis, which states that the extents of sharedness of SSR are different for the types of conclusions. To test which pairs significantly differed, the *post hoc* Games-Howell test is conducted. The test results indicted a statistically significant difference between 'missing conclusions' and 'explicit conclusions' (p = .01), and between 'missing conclusions', and 'implicit conclusions' (p = .04). The estimated omega squared (ω^2 = .04), indicated that approximately 4% of the total variation in average extents of sharedness of SSR is attributable to differences between the three types of conclusions. This is considered a small effect size.

Second, the descriptive statistics with respect to the operationalization 'Relative frequency HLTC' (B) (Table 7), demonstrated that 'explicit conclusions' show the highest median on the extent of sharedness

of SSR (*Mnd* = .60), whereas 'missing conclusions' show the lowest median on the extent of sharedness of SSR (*Mnd* = .50). Before analysis, the assumptions of normality, and equality of variances are examined. The normality check, using skewness and kurtosis (Table 7) indicated non-normal distribution of the data. The variances are compared using Levene's test. The test result indicated significantly heterogeneous variances (F = 2, 210) = 6.12, p = .003). As the data violated both the assumption of homogeneity of variances, and normality, the non-parametric Kruskal-Wallis H test is selected. There is, however, an important assumption that needs to be tested prior to the Kruskal-Wallis H test application. It requires that the data in each group (the teams) have the same distributions, and thus the same variability (Kruskal & Wallis, 1952). Unlike the traditional Levene's test, the non-parametric equivalent tests for the same distributions. The test result is non-significant $F(2) = 2.319$, $p = .101$, which enabled us to apply the Kruskal-Wallis H test. It showed a statistically non-significant test result $\chi^2(2) = 4.643$, $p = .098$ (Table 8). Hence, we confirm the null hypothesis, and assume that the extents of sharedness of SSR are not different for the various conclusions.

Third, the descriptive statistics concerning operationalization 'Relative duration CTM' (C) (Table 6), deduced that 'explicit conclusions' have the numerically highest median extent of sharedness of SSR (*Mnd* = .14), whereas 'implicit conclusions' have the smallest median (*Mnd* = .12). Prior to the analysis, the assumption of normality, and homogeneity of variances are checked. The normality test using kurtosis and skewness, indicated that the distribution of the data is non-normal. A Levene's test based on the non-parametric equivalent, showed non-significant variances ($F(2, 210) = 1.640$, $p = .196$). As the data violated the assumption of normality, we decided to use the non-parametric Kruskal Wallis test. It revealed a statistically non-significant test result $\chi^2(2) = 4.056$, $p = .132$ (Table 8). Thus, we accept the null hypothesis of no differences in the extents of sharedness of SSR between the types of conclusions.

Fourth, the descriptive data for the operationalization 'Relative number CTM' (D) (Table 7), show that episodes that contained 'explicit conclusions' are accompanied with the numerically highest mean extent of sharedness of SSR (M = .69), and episodes with 'missing conclusions' with the lowest mean extent (M = .54). Prior to the analysis, the assumption of normality, and homogeneity of variances it investigated. The normality assumption is checked with skew and kurtosis. Normality distribution is accepted, as the skew and kurtosis z-scores (Table 7) did not exceed the critical value of ±3.29 for medium-sized samples ($50 < n < 300$) (Kim, 2013). The variances are compared with Levene's test The test statistic demonstrated a significant test result $F(2, 210) = 3.30$, $p = .039$, therefore we did assume unequal variances. Since the assumption of homogeneity of variances is rejected, we performed a Welch's test. It showed a statistically significant difference in extents of sharedness of SSR between the various types of

conclusions, $F(2, 82.91) = 6.123$, $p < .003$, so we rejected the null hypothesis of no differences between the extents of 'shared conclusion-drawing'. To establish which pairs significantly differed, the Games-Howell *post hoc* test is carried out. The test results only suggested a statistically significant difference between 'missing conclusions', and 'explicit conclusions' ($p = .001$). Additionally, we calculated that within the 'Relative number CTM' (D) operationalization, the extent of 'shared conclusion-drawing' accounted for a small size of 5% ($\omega^2 = .05$) of the total variance related to the extents of sharedness of SSR.

Concluding, the extents of sharedness of SSR are statistically significantly related to higher extents of 'shared conclusion-drawing', for the operationalizations 'Relative duration HLTC' (A) and 'Relative number CTM' (D). This means that we can confirm hypothesis 1: high extent of sharedness of SSR is positively related to high extent shared conclusion-drawing.

Testing Hypothesis 2. To test hypothesis 2, and examine the external validity of our results, SQ b is addressed.

SQ b: Can variations in the extents of sharedness of SSR be attributed to the 'type of Scrum meeting' ('daily Scrum', and 'retrospective')?

For that purpose we performed a series of independent sample t-tests, and a Mann–Whitney U-test, with $\alpha = .05$ as a threshold for statistical significance. The independent variables represented the two types of Scrum meetings (daily Scrum, retrospective). The dependent variable are the extents of sharedness of SSR detected by the separate operationalizations. The normality assumption is examined using skew and kurtosis (e.g., Mardia, 1970). As can be seen from Table 9, expect for operationalization 'Relative duration CTM' (C), the data are sufficiently normally distributed for the purpose of conducting t-test (i.e., z-scores skewness, and kurtosis < ±3.29) (Kim, 2013). Table 9 summarizes, the means, standard, normality test, and t-test results on the extents of sharedness of SSR per operationalization.

First, the descriptive statistics of operationalization 'Relative duration HLTC' (A) show that the mean extents of sharedness of SSR are numerically higher for 'retrospectives' ($M = .54$) than for 'daily Scrums' ($M = .53$). Variances are compared, using Levene's test. The result demonstrated unequal variances ($F = 7, 829$, $p = .006$), so the degrees of freedom are adjusted from 211 to 173,174. An independent-samples t-test demonstrated no statistically significant effect of the 'type of Scrum meeting' on the average extents of sharedness of SSR between 'daily Scrum' ($M = .53$, $SD = .33$) and 'retrospectives' ($M = .54$, $SD = .39$), $t(173.174) = -0.22$, $p = .83$. Thus, we confirm the null hypothesis, and assume no statistically significant difference in the extents of sharedness of SSR between 'daily

Scrums', and 'retrospectives'. We can be 95% confident that the true difference between these means is between CI = [-0.11059; 0.08865].

Second, as for operationalization 'Relative frequency HLTC' (B), the descriptive statistics show that the mean extents of sharedness of SSR are numerically higher for 'retrospectives (M = .50) than for 'daily Scrums' (M = .50). Variances are compared with Levene's test. The result showed heterogeneous variances (F = 9, 807, p = .002), so degrees of freedom are adjusted from 211 to 166,612. An independent-samples t-test revealed a non-significant difference in the difference in the extents of sharedness of SSR between 'daily Scrums' (M = .50, SD = .28) and 'retrospectives' (M = .50, SD = .34), t(166.612) = -0.14, p = 0.89. Therefore, we accept the null hypothesis, and conclude that there is no significant difference in the extents of sharedness of SSR, between 'daily Scrums', and 'retrospectives'. We can assert with 95% confidence that the true difference between these means is between CI = [-0.09351; 0.08103].

Third, the descriptives regarding the operationalization 'Relative duration CTM' (C) reveal that the median extent of sharedness of SSR is higher for 'daily Scrums' (*Mnd* = .16) than for 'retrospectives' (*Mnd* = .13). The variances are checked with the Levene's test based on the non-parametric equivalent. It showed homogeneous variances (F(2, 210) = 1.640, p = .196). This allowed us to apply the Mann–Whitney U-test is non-significant at the p = .08 (Table 9). Which leads us to accept the null hypothesis, and conclude that the extents of sharedness of SSR are not significantly different for 'daily Scrums', than for 'retrospectives'. We can be 95% confident that the true difference between these means is between CI = [0.04920; 0.20721].

Fourth, the descriptive data concerning the 'Relative number CTM' (D) operationalization, show that the average extents of sharedness of SSR are numerically higher for 'daily Scrums' (M = .69) than for 'retrospectives' (M = .55). Variances are compared using Levene's test. It revealed unequal variances (F = 6, 474, p = .012), so degrees of freedom are adjusted from 211 to 168,653. An independent-samples t-test is significant: t(168.653) = 3.68, p < 0.01 (Table 9). Which causes us to reject the null hypothesis, and assume that the extents of sharedness of SSR are statistically significantly higher for 'daily Scrums', than for 'retrospectives. To measure the effect size of the 'type of Scrum meetings' on the extents of sharedness of SSR, Cohen's d is calculated. Values between .0 and .20 indicate a negligible effect size (Cohen, 1988). Cohen's d = 0.05, which thus indicated a negligible practical effect. We can conclude with 95% confidence that the true extent between 0.06226 and 0.20622.

In sum, the operationalization 'Relative number CTM' (D) only found a statically significant difference in the extents of sharedness of SSR between 'the meeting types'. Therefore, hypothesis 2 is rejected: differences in the extent of sharedness of SSR are attributable to the 'type of Scrum meeting' ('daily Scrums', 'retrospectives').

Table 9

Summary of Means, and Standard Deviations for Extents of Sharedness of SSR for Daily Scrums and Retrospectives

Operationalization	Meeting Type	*N*	*M*	*Mnd*	*SD*	Skewness	*Z Skew*	*Kurtosis*	Z Kurtosis	*T*	*U*	*p* <	η^2	Cohen's *d*
Rd HLTC (A)	Overall	213								-0.22	-	.83	-	.03
	DS	123	.53	.64	.33	-0.548	-2.51	-1.149	-2.65					
	R	90	.54	.68	.39	-0.455	-1.79	-1.496	-2.97					
Rf HLTC (B)	Overall	213								-0.14	-	.89	-	.02
	DS	123	.50	.57	.28	-0.789	-3.62*	-0.595	-1.37					
	R	90	.50	.60	.35	-0.554	-2.18	-1.308	2.60					
Rd CTM (C)	Overall	213								-	4755	.08	.015	-
	DS	123	.24	0.16	.25	1.235	5.66*	0.703	1.62					
	R	90	.19	0.13	.22	1.353	5.33*	1.303	2.59					
Rn CTM (D)	Overall	213								3.68	-	< .01	-	.05
	DS	123	.69	.67	.23	0.075	0.34	-1.132	-2.61					
	R	90	.55	.50	.28	0.271	1.07	-1.141	2.27					

Note. 0.05 significance level is used. T-test statistics, and Cohen's d effect sizes are displayed. *Z-scores that exceeded the critical z-score ±3.29 for normality. For operationalization 'Relative frequency HLTC' this is not a serious problem for t-tests given that these do not necessarily require normal distribution (Lumley, Diehr, Emerson, & Chen, 2002). DS = Daily Scrum; R = Retrospective; Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members.

4. Discussion

The purpose of the present study was to explore how the extents of sharedness of Socially Shared Regulation (SSR) can be determined. For this, two measurement methods are developed; one with a qualitative approach, and one with a quantitative approach. This study examined which method (or combination) is most appropriate to serve as a measure for the extent of sharedness of SSR. Both methods are found suitable measures. In this section, these results will be discussed in detail. After that, the theoretical implications, directions for future research, study limitations, and practical implications are examined. The thesis concludes with a conclusion.

4.1 Discussion of the Results

This exploratory study answered the following main research question.

RQ: How can the extent of sharedness of SSR in Scrum teams be determined?

Findings from the present study indicate that a combination of the two presented methods would generate the most satisfying results regarding the extents of sharedness of SSR. This implies a combination of one method with a qualitative approach, that operationalizes the extent of sharedness of SSR based on the ‘frequencies’, and ‘durations’ of ‘high-level transactive communication’; complemented with one method with a quantitative approach that operationalizes the extents of sharedenss of SSR based on the ‘number’ of team members involved in team conversations. This selection is made based on five quality criteria, that is, that they are theoretical substantive, appropriate to perform measures on coded video data, sufficiently discriminative, and easy-to-use. The decisive factor for the appropriateness of the measurement method, is that the extents of sharedness of SSR detected by the various operationalizations are sufficiently similar to those of one or more other operationalizations.

Nevertheless, one operationalization of the method with the quantitative approach, that operationalizes the extents of sharedness of SSR based on the ‘durations’ of team member contributions seems insufficient as a measure; as this operationalization did not detect similar extents of sharedness of SSR as the other operationalizations. A possible explanation is that this operationalization detects many *extreme* extents of sharedness of SSR. For instance, if one person speaks during an entire episode, the extent of sharedness of SSR for that episode is zero. It should, however, be noted here that the extreme extents detected in this dataset, might not be representative for Scrum teams in general. Therefore, we do not reject this operationalization as possible measure, and include it in the discussion of the results.

Discussing the methods and operationalizations. On the basis of the results we can conclude that the methods, and therefore the operationalizations, have a number of advantages, and disadvantages. We start with the discussion of the advantages, after which the disavantages are discussed.

Advantages. Advantages that apply for all four operationalizations is their easy application to determine the extents of sharedness of SSR. This applies, above all, to the operationalizations 'Relative duration CTM' (C) and 'Relative number CTM' (D, of method 2); these operationalizations calculate the extents of sharedness of SSR by mere 'statistical calculations'. Data-analysis using the method with the qualitative approach (based on 'high-level transactive communication') requires more time, and effort; as the video-data need to be analysed first, using a 'coding scheme'. An additional advantage is that the measurement method had a different approach (qualitative, quantitative). Additionally, each method had two operationalizations that focussed on a different aspect of sharedness of SSR. For instance, the method with the qualitative approach focused on the extent to which SSR is shared, based on the extent to which 'high-level transactive communication' is used; one operationalization focused on the 'durations' (Relative durations HLTC', A), and one focused on the 'frequencies' ('Relative frequency HLTC', B). The same holds for the operationalizations of the method with the quantitative approach (based on 'equal team member contributions'); one focused on the 'durations' ('Relative durations CTM', C) whereas the other focused on the 'frequencies' ('Relative number CTM', D). In this respect, it can be argued that the complementary nature of these operationalizations yields a comprehensive picture of the extent to which SSR is shared among team members. However, these four operationalizations are also accompanied by a number of drawbacks.

Disadvantages. The disadvantages are related to both the application of all four operationalizations, and to application-specific limitations of the separate operationalizations. We start this discussion with the general limitations of the operationalizations. The general limitations relate to contrasting findings; which are illustrated in Box 1, 2, 3 and 4 on the following two pages. There a eight examples of contrasting extents of sharedness of SSR possible between the two methods.

Box 1.

Contrasting Scenarios Operationalization 'Relative duration HLTC' (A) versus 'Relative duration CTM' (C)

Contrasting scenario 1:

1. the 'duration' of transactive communication is 0 seconds for 'ignoring'; 350 seconds for 'accepting'; respectively **150** seconds for 'engaging' (total duration is **500** seconds).
2. the shortest speaking 'duration' is **200** seconds, and the longest speaking duration is **250** seconds.

1. **Extent of sharedness of SSR** of operationalization 'Relative duration HLTC' (A) is **150** / **500** = .3
2. **Extent of sharedness of SSR** of operationalization 'Relative duration CTM' (C) is **200** / **250** = .8

Contrasting scenario 2:

1. the 'duration' of transactive communication is 0 seconds for 'ignoring'; 150 seconds for 'accepting'; respectively **350** seconds for 'engaging' (total duration is **500** seconds).
2. the shortest speaking 'duration' is **50** seconds, and the longest speaking duration is **250** seconds.

1. **Extent of sharedness of SSR** of operationalization 'Relative duration HLTC'(A) is **350** / **500** = .7
2. **Extent of sharedness of SSR** of operationalization 'Relative duration CTM'(C) is **50** / **250** = .2

Conclusion based on scenario 1 and 2: contrasting extents of sharedness of SSR are detected when either team members use low 'frequencies' of 'engaging' communication ('Relative duration HLTC' is low), and the difference between the longest and the shortest speaking duration is small ('Relative duration CTM' is high), and vise versa.

Box 2.

Contrasting Scenarios Operationalization 'Relative frequency HLTC' (B) versus 'Relative number CTM' (D)

Contrasting scenario 3:

1. the 'frequencies' of transactive communication are 0 for 'ignoring'; 20 for 'accepting'; respectively **5** for 'engaging' (total frequency is **25**).
2. the 'number' of team members involved in topic is **5**, the 'number' of team members present is **5**.

1. **Extent of sharedness of SSR** of operationalization 'Relative frequency HLTC' (B) is **5** / **25** = .2
2. **Extent of sharedness of SSR** of operationalization 'Relative number CTM'(D) is **5** / **5** = 1

Contrasting scenario 4:

1. the 'frequencies' of transactive communication are 0 for 'ignoring'; 5 for 'accepting'; respectively **20** for 'engaging' (total frequency is **25**).
2. the 'number' of team members involved in topic is **1**, and the 'number' of team members present is **5**.

1. **Extent of sharedness of SSR** of operationalization 'Relative frequency HLTC'(B) is **20** / **25** = .8
2. **Extent of sharedness of SSR** of operationalization 'Relative number CTM' (D) is 1 team member engaged is 0 according to the exception.

Conclusion based on scenario 3 and 4: contrasting extents of sharedness of SSR are detected when either team members use low 'durations' of 'engaging' communication ('Relative frequency HLTC' is low), and the difference between the 'number' of team members involved in topic is high ('Relative number CTM' is high), and vise versa.

Box 3.

Contrasting Scenarios Operationalization 'Relative duration HLTC' (A) versus 'Relative number CTM' (D)

Contrasting scenario 5:

1. the 'duration' of transactive communication is 0 seconds for 'ignoring'; 350 seconds for 'accepting'; **150** seconds for 'engaging' (total duration is **500** seconds).
2. the 'number' of team members involved in topic is **5**, and the 'number' of team members present is **5**.

1. **Extent of sharedness of SSR** of operationalization 'Relative duration HLTC' (A) is **150** / **500** = .3
2. **Extent of sharedness of SSR** of operationalization 'Relative number CTM' (D) is **5** / **5** = 1

Contrasting scenario 6:

1. the 'duration' of transactive communication is 0 seconds for 'ignoring'; 150 seconds for 'accepting'; **350** seconds for 'engaging' (total duration is **500** seconds).
2. the 'number' of team members involved in topic is **2**, and the 'number' of team members present is **5**.

1. **Extent of sharedness of SSR** of operationalization 'Relative duration HLTC' (A) is **350** / **500** = .7
2. **Extent of sharedness of SSR** of operationalization 'Relative number CTM' (D) is **2** / **5** = .4

Conclusion based on scenario 5 and 6: contrasting extents of sharedness of SSR are detected when either team members use low 'durations' of 'engaging' communication ('Relative duration HLTC' is low), and the 'number' of team members involved in topic is high ('Relative duration CTM' is high), and vise versa.

Box 4.

Contrasting Scenarios Operationalization 'Relative frequency HLTC' (B) versus 'Relative duration CTM' (C)

Contrasting scenario 7:

1. the 'frequencies' of transactive communication are 0 for 'ignoring'; 20 for 'accepting'; respectively **5** for 'engaging' (total frequency is **25**).
2. the shortest speaking 'duration' is **200** seconds, and the longest speaking duration is **250** seconds.

1. **Extent of sharedness of SSR** of operationalization 'Relative frequency HLTC' (B) is **5** / **25** = .2
2. **Extent of sharedness of SSR** of operationalization 'Relative duration CTM'(C) is **200** / **250** = .8

Contrasting scenario 8:

1. the 'frequencies' of transactive communication are 0 for 'ignoring'; 5 for 'accepting'; respectively **20** for 'engaging' (total frequency is **25**).
2. the shortest speaking 'duration' is 50 seconds, and the longest speaking duration is **250** seconds.

1. **Extent of sharedness of SSR** of operationalization 'Relative frequency HLTC' (B) is **20** / **25** = .8
2. **Extent of sharedness of SSR** of of operationalization 'Relative duration CTM' (C) is **50** / **250** = .2

Conclusion based on scenario 7 and 8: contrasting extents of sharedness of SSR are detected when either team members use low extent of 'engaging' communication ('Relative duration HLTC' is low), and the difference between the longest and the shortest speaking duration is large ('Relative duration CTM' is low), and vise versa.

Now, the application-specific limitations of the separate operationalizations are discussed. A disadvantage of operationalization 'Relative duration HLTC' (A) is that the difference in 'durations' between the favorable 'accepting' utterances, and the unfavorable 'ignoring' utterances is not reflected in the extent of sharedness of SSR (see Box 5; scenario 1, and 2). As a consequence, a distorted picture can emerge of the extent to which SSR is shared.

Box. 5

Disadvantage Operationalization 'Relative duration HLTC' (A)

$$Extent\ of\ sharedness\ of\ SRR = \boldsymbol{Rd\ HLTC} = \frac{Time\ of\ engaging\ utterances}{Total\ time\ of\ regulation\ utterances}$$
Disadvantage: operationalization does not distinct between 'ignoring' and 'accepting' utterances
Scenario 1: the duration of **'**accepting' utterances is long, relative to 'ignoring' utterances. TM a: 'ignoring': 0 seconds; 'accepting': **70** seconds; 'engaging': 0 seconds. TM b: 'ignoring': 0 seconds; 'accepting': **20** seconds; 'engaging': **400** seconds. TM c: 'ignoring': 0 seconds; 'accepting': **10** seconds; 'engaging': **200** seconds. **Extent of sharedness of SSR = 600** / **700** = .86
Scenario 2: the duration of **'**ignoring' utterances is long, relative to 'accepting' utterances. TM a: 'ignoring': **70** seconds; 'accepting': 0 seconds; 'engaging': 0 seconds. TM b: 'ignoring': **20** seconds; 'accepting': 0 seconds; 'engaging': **400** seconds. TM c: 'ignoring': **10** seconds; 'accepting': 0 seconds; 'engaging': **200** seconds. **Extent of sharedness of SSR = 600** / **700** = .86
Finding scenario 1 and 2: the extents of sharedness of SSR are the same for scenario 1, and 2 (.86) despite that the durations of 'ignoring' and 'accepting' utterances differs.
Conclusion: a disadvantage is that differences between the durations of 'ignoring' and 'accepting' utterances are not reflected in the extent of sharedness of SSR.

A disadvantage of operationalization 'Relative frequency HLTC' (B) is that difference in 'frequency' of the favourable 'accepting' utterances, and the unfavourable 'ignoring' utterances is not reflected in the extent of sharedness of SSR (see Box 6; scenario 3, and 4). As a result, a distorted picture can emerge of the extent of sharedness of SSR. An additional disadvantage of this operationalization is that it does not reflect how many team members provide a constructive contribution; use 'engaging' communication (see Box 6; scenario 5, and 6). Therefore, it is not possible to see how many team members use less 'engaging' communication.

Box. 6

Disadvantages Operationalization 'Rf HLTC' (B)

$Extent\ of\ sharedness\ of\ SRR = \boldsymbol{Rf\ HLTC} = \frac{Frequency\ of\ engaging\ utterances}{Total\ frequency\ of\ regulation\ utterances}$
Disadvantage: operationalization does not distinct between 'ignoring' and 'accepting' utterances
Scenario 3: relative high 'frequency' of **'**accepting' utterances, compared to 'ignoring' utterances. TM a: 'ignoring': **15**; 'accepting': **5**; 'engaging': 0. TM b: 'ignoring': 0; 'accepting': **10**; 'engaging': **25**. TM c: 'ignoring': 0; 'accepting': **5**; 'engaging': **20**. **Extent of sharedness of SSR = 45** / 80 = .56
Scenario 4: relative low 'frequency' of **'**accepting' utterances, compared to 'ignoring' utterances. TM a: 'ignoring': 0; 'accepting': **20**; 'engaging': 0. TM b: 'ignoring': 0; 'accepting': **10**; 'engaging': **25**. TM c: 'ignoring': 0; 'accepting': **5**; 'engaging': **20**. **Extent of sharedness of SSR = 45** / 80 = .56
Finding scenario 3 and 4: the extents of sharedness of SSR are the same for scenario 1, and 2 (.56); despite that the 'frequency' of 'ignoring', and 'accepting' utterances differs.
Conclusion: a disadvantage of this operationalization is that the difference in 'frequency' between 'ignoring', and 'accepting' utterances is not reflected in the extent of sharedness of SSR.
Disadvantage: individual contributions are not reflected
Scenario 5: 2 team member use 'engaging' communication, while one other does not. TM a: 'ignoring': 0; 'accepting': 0; 'engaging': **4**. TM b: 'ignoring': 0; 'accepting': 0; 'engaging': **1**. TM c: 'ignoring': 0; 'accepting': 0; 'engaging': 0. **Extent of sharedness of SSR** = 5 / 5 = 1
Episode scenario 6: all 3 team members use 'engaging' communication. TM a: 'ignoring': 0; 'accepting': 0; 'engaging': **2**. TM b: 'ignoring': 0; 'accepting': 0; 'engaging': **2**. TM c: 'ignoring': 0; 'accepting': 0; 'engaging': **1**. **Extent of sharedness of SSR** = 5 / 5 = 1
Finding scenario 5 and 6: the extent of sharedness of SSR is the same for scenario 1, and 2. This number is thus not affected by the 'frequency' of 'engaging' utterances across the team members.
Conclusion: the extent of sharedness of SSR does not reflect the 'frequency' of 'engaging' utterances across the team members. Therefore, it is not clear how many team members do provide constructive contributions (use 'engaging' communication). A disadvantage of this operationalization is thus that the extents of sharedness of individual contributions are not reflected in the outcome.

A disadvantage of operationalization 'Relative duration CTM' (C) is that it only takes into calculations the extreme speaking durations; the sum of the shortest, and the longest speaking duration (see Box 7; scenario 7, and 8). Therefore, the extent of sharedness of SSR does not represent how the speaking durations are spread across the team members. Consequently, if the extent of sharedness of SSR is rather low the data on the distributions of team member contributions need to be consulted; to see where improvement can be made.

Box 7.

Disadvantage Operationalization 'Relative duration CTM (C)

$Extent\ of\ sharedness\ of\ SRR = \boldsymbol{Rd\ CTM} = \frac{Sum\ shortest\ speaking\ duration\ (episode)}{Sum\ longest\ speaking\ time}$
Disadvantage: operationalization only takes into account the extremes in speaking durations
Scenario 7: the sum of the intervening speaking durations are close to the sum of the shortest speaking duration. (Sum of speaking durations per team member: TMa: **15** sec.; TMb: 25 sec.; TMc: 30 sec.; TMd: 30 sec.; TMe: **600** sec.). **Extent of sharedness of SSR = 15** / **600** = .03
Scenario 8: the sum of the intervening speaking durations are close to the sum of the longest speaking duration. (Sum of speaking durations per team member: TMa: **15** sec.; TMb: 500 sec.; TMc: 500 sec.; TMd: 550 sec.; TMe: **600** sec.). **Extent of sharedness of SSR = 15** / **600** = .03
Finding scenario 7 and 8: the extent of sharedness of SSR is the same for scenario 1, and 2. Thus, the variations in the sum of the intervening speaking durations do not affect the extent of sharedness of SSR.
Conclusion: a disadvantage of this operationalization is that the extent of sharedness of SSR is only affected by the sum of the shortest, and the longest speaking duration.

A disadvantage of operationalization 'Relative number CTM' (D) is that the extent of sharedness of SSR of a small 'number' of team members present at the meeting is stronger affected by the 'number' of team members involved in topic, than in large teams (see Box 8; scenario 9, 10, 11, and 12). A further disadvantage of this operationalization is that limited variation in the extents of sharedness of SSR is possible (see Box 8; scenarios 13 and 14). The smaller the team, the larger these disadvantages become. An implication is therefore that, it becomes more difficult to compare the extents of sharedness of SSR between teams of different sizes.

Box 8.

Disadvantages operationalization 'Relative number CTM (D)

$Extent\ of\ sharedness\ of\ SRR = \boldsymbol{Rn\ CTM} = \frac{Numer\ of\ team\ members\ involved\ in\ topic}{Total\ number\ of\ team\ members\ present}$
Disadvantage: operationalization does not account for different team sizes
Scenario 9: **1** team member is involved, at a meeting where **6** members are present. **Extent of sharedness of SSR** = **1** / **6** = 0.17
Scenario 10: **5** team members are involved, at a meeting where **6** members are present. **Extent of sharedness of SSR** = **5** / **6** = 0.83
Scenario 11: **1** team member is involved, at a meeting where **3** members are present. **Extent of sharedness of SSR** = **1** / **3** = 0.33
Scenario 12: **2** team members are involved, at a meeting where **3** members are present. **Extent of sharedness of SSR** = **2** / **3** = 0.67
Finding scenario 9 and 10: when a large 'number' of team members is present at the meeting, the extent of sharedness of SSR is less affected by the 'number' of team members involved in topic; compared to small teams.
Finding scenario 11 and 12: when a small 'number' of team members is present at the meeting, the extent of sharedness of SSR is more affected by the 'number' of team members involved in topic; compared to large teams.
Conclusion: the less team members are involved in the team meetings, the stronger the extent of sharedness of SSR is affected by the 'number' of team members involved in topic. A disadvantage of this operationalization is thus that the extents of sharedness of SSR are not equally accounted for by different team sizes.
Disadvantage: limited variation is possible
Scenario 13: six team members are present at the meeting. **Extent of sharedness of SSR** can only be either 0, 17, 33, .50, 67, 83, or 1.
Scenario 14: three team members are present at the meeting. **Extent of sharedness of SSR** can only be either 0, 33, 67, or 1.
Finding 13: when a large 'number' of team member is present at the meeting, more variation in the extent of sharedness of SSR is possible; compared to small teams.
Finding 14: when a small 'number' of team member is present at the team meeting, little variation in the extent of sharedness of SSR is possible; compared to large teams.
Conclusion: the less team members are present at the meeting, the less variation in the extents of sharedness of SSR is possible. A disadvantage of this operationalization is thus that it allows for limited varation in the extents of sharedness of SSR.

In sum, the discussion of the advantages and disadvantages of the two measurement methods, and their four operationalizations shows that the operationalization are somewhat competing, but are also complementary and easy to use examinations of the extents of sharedness of SSR.

To explain variations in the extents of sharedness of SSR, we explored whether variations are related to the extents of 'shared conclusion-drawing'. This is addressed by sub question a.

SQ a: Can high extents of sharedness of SSR be attributed to high extents shared conclusion-drawing (missing conclusion, implicit conclusion, and explicit conclusion)?

The results of both the qualitative, and quantitative measurement method show that SSR is more shared when high-extent shared conclusions (either 'implicit', or explicit conclusions') are used. This appears to apply in particular to 'explicit conclusion-drawing'. This finding is consistent with previous studies, and our prior expectations on 'explicit conclusion-drawing', to be positively associated with high extents of sharedness of SSR (e.g., van der Haar et al., 2013). Concluding, is seems that 'conclusion-drawing' positively attributes to high-extent sharedness of SSR.

When examining the differences between 'implicit', and 'explicit conclusion-drawing', no operationalization identified relevant differences in the extents of sharedness of SSR. This finding is neither consistent with Raes et al.'s (2015) theoretical model, nor with our expectations on 'explicit conclusion-drawing'. It might be possible that our distinction between 'implicit conclusions', and 'explicit conclusions' is not clearly formulated. In that case, future studies should focus on attempts to develop a more fine-grained distinction between these types of conclusions. Another, explanation is that the number of teams in our sample is too limited to draw conclusions from.

In conclusion, based on our findings and considering earlier findings, the extent of 'shared conclusion-drawing' seems an appropriate aspect of SSR to explain variations in the extents of sharedness of SSR.

We also investigated to what extent variations in the extents of sharedness of SSR are related to the 'type of Scrum meeting'; for that purpose we addressed sub question b.

SQ b: Can variations in the extents of sharedness of SSR be attributed to the 'type of Scrum meeting' ('daily Scrum', and 'retrospective')?

The results revealed that variations could not be attributed to the 'type of Scrum meeting'. The method with the qualitative approach (based on 'high-level transactive' communication) did not find different extents of sharedness of SSR between 'daily Scrums' and 'retrospectives'. The same holds for the method 'equal team member contributions', that is operationalized based on the durations of team

member contributions. From this it can be argued that team members do not more engage in or related to each other's SR processes, during 'daily Scrums' than in 'retrospectives'. Contradictory, the results of the method 'equal team member contributions', that operationalizes the extent of sharedness of SSR based on the number of contributing team members, did reveal higher extents of sharedness of SSR for 'daily Scrums' than for 'retrospectives'. A possible explanation is that team members manage team member involvement more effectively during 'daily Scrums' than during 'retrospectives' because 'daily Scrums' are subjected to time constraints (Dybå, Dingsøyr, & Moe, 2014).

Overall, three of the four operationalizations did not explain variations in the extents of sharedness of SSR, based on the 'type of Scrum meeting'. According the quality criteria, this aspect of collaboration thus seems inadequate to explain variations in the extents of sharedness of SSR. However, we found some evidence indicating that the extents of sharedness of SSR are higher for 'daily Scrums', than for 'retrospectives'.

4.2 Theoretical Implications and Directions for Future Research

This study contributes to existing literature on SR. The findings expand current knowledge on, and understanding of measures for the extent to which SSR is shared among team members. The results have several theoretical implications that are addressed briefly.

A possible research opportunity is to *expand our research* findings. A particularly interesting topic for future study would be to assess additional literature-based aspects which could be a valid measure for the extent of sharedness of SSR. In literature, several communication-related aspects are described that could serve as measurement variables for the extent to which SSR is shared. Possible aspects that could be considered are *goal-focused talk* (Volet et al., 2009), *fundamental* versus *surface aspects* of regulation (Grau & Whitebread, 2012), or taking in a *I*, *You*, or *We* perspective (Schoor et al., 2015). In addition, including multiple aspects is necessary to yield a more reliable measure for the extent of sharedness. Therefore, a multimethod-approach, like ours, is prevailed.

Another research opportunity is to investigate the *generalizability* of the measurement methods (and operationalizations) in other settings. Future studies should be conducted to determine whether the findings are similar in other settings. For example, our research could be replicated in educational settings, or in organizational settings (other than Scrum teams) that depend on self-directed teams. The findings of such studies could indicate the necessity to consider additional context-specific aspects, when using the extent to which SSR is shared to determine effective SR.

Future studies could also focus on investigating the *cost benefits for organizations* regarding

measures for the extent of sharedness of SSR. Carrying out the measurement analysis for the extent to which teams share SSR, can be accompanied by high costs. Therefore, it would be interesting to conduct studies looking for a set of indicators that determine the necessity for such measurement controls.

A final research opportunity, is to combine measuring interaction-related aspects, with *non-verbal* aspects to measure the extents of sharedness of SSR. Our study focussed on analysing verbal expressions of team interaction in meetings. Literature, however, also highlights the importance of including *non-verbal* expressions (e.g., mimicry, body language) when assessing effective SR (Saarni, 1982). Research in this area is, however, rather limited. For future studies, it is recommended to also include *non-verbal* aspects to determine the extents of sharedness of SSR.

4.3 Limitations and Directions for Future Research

As each study, ours is accompanied by a few limitations. These are related to the study design, and methodological limitations.

First, the *generalizability* of the results could be improved. This study empirically tested two measurement methods on a limited sample of Scrum teams. Future studies should include a larger sample size to draw generalizations about the extent to which SSR is shared in Scrum teams. The current sample is too small to provide a definitive answer to the question of the extent to which SSR is shared among team members. However, the extents to which SSR is shared among Scrum team members is not the main interest of this current study. Therefore, this limitation is of minor importance to serve our study purpose.

Second, *methodological improvement* can be achieved. Especially with regard to measures on the extents of sharedness of SSR, which required development of measurement methods. Although several reliability, and validity strategies are employed, further analysis by other studies is needed. For instance, by adding additional data resources, such as documents related to the decisions made at meetings. This might provide valuable insights into the extents of sharedness of regulation processes of Scrum teams. For instance, the agendas of the meetings can be consulted to assess whether agreements are noted, and thus shared or not. In this way, the team members, also the ones that did not attend the meeting, can have easily access to the agreements made. This limitation is not of crucial importance for this present study, but can provide can provide a more comprehensive measurement approach.

Third, further *validation* of the measurement methods, and operationalizations is required. This research gives a first impetus towards measuring the extents to which SSR is shared, based on the measurement methods presented here. This implicates that we should interpret our results cautiously. Logically, the measurement methods need further validation in similar settings to provide a definite answer

to the question of how to determine the extents of sharedness of SSR. Therefore, as with all exploratory studies, additional studies are required to validate the methods (Baarda, 2010). For instance, replication studies could be conducted with (more) video-data of other Scrum teams that are employed at other Software Development Organizations. This limitation is not a major shortcoming of this present study; as this is a common limitation for exploratory study designs.

Fourth, this study attempted to explain variations in the extents of sharedness of SSR based on the 'type of Scrum meeting'. As aforementioned, we had *insufficient data* on 'sprint planning meetings' to analyse for the extents of sharedness of SSR. Additionally, an additional limitation is that the data on the 'sprint review meetings' were not appropriate for data-analysis because external parties were involved. Therefore, a study implication is that variations in extents of sharedness of SSR could not be explained across all 'types of Scrum meetings'. Perhaps future studies could collect data on all four 'types of Scrum meetings' ('sprint planning meeting', 'daily Scrum', 'sprint review meeting', 'retrospective'), and make an attempt to explain variations in the extents of sharedness of SSR. This limitation confined this study to only make statements on the differences in extents of sharednesss of SSR between 'daily Scrums' and 'retrospectives'. It is expected that future studies, will demonstrate differences in the extents of sharedness of SSR between the four 'types of Scrum meetings'. This expectation follows from the fact that each 'type of Scrum meeting' has its own purpose, and duration.

Fifth, this present study did not assess the *substantive validity* of the verbal interactions. In the analysis, one of our assumptions is that the involved software engineers are considered experts within their domain of knowledge. We did not investigate whether their statements are correct in terms of content. This implicates that we can not draw any conclusions about the discussed domain-specific knowledge. Consequently, for future research, it is advisable to benefit from expert input from those in the field of software engineering. By doing this, it is expected that a more insightful view can be obtained on the extent to which SSR is 'correctly' shared. A possible area for future investigation would be the extent to which team members correct each other's inaccuracies in reasoning; with respect to correctness of content-related aspect of shared regulatory processes. For this present study, this limitation is not as serious as it would be when investigating mere content-related SSR.

Sixth, the coding scheme 'shared conclusion-drawing' presented needs *further refined*. This coding scheme is developed to code the data for the extent of 'shared conclusion-drawing'. The data revealed differences in the extents of sharedness of SSR between the conclusions. Against expectations, not all categories are significantly distinct in the detected extents of sharedness of SSR. Possibly, the coding categories should be reconsidered. This implied that we were not able to make convincing statements of

the relationship between the ‘the extent of ‘shared conclusion-drawing’, and the extents of sharedness of SSR. Based on current literature, and our expectations it is predicted that improved coding categories can find significant differences in the levels of sharedness between the three ‘types of conclusions’.

4.4 Practical Implications

The findings of the current study may have implications for organizational decisions, and training programs on effective SR; of self-directed (Scrum) teams.

First, organization managers need to *consider training needs* in approaching shared interaction-related aspects of their self-directed teams. The findings of our study may reflect the need for organization managers to identify and/or map interaction-related challenges. Some examples of these challenges are effective meeting running (Cho, 2008), shared conclusion-drawing (e.g., Rogat & Linnenbrink-Garcia, 2011; van der Haar et al., 2013), and equal team member contributions (e.g., Hoegl & Gemuenden, 2001; Volet et al., 2009). To this end, managers should consider undertaking training investments tailored to stimulate the sharedness of SSR (and thus effective SSR) of Scrum teams. Through needs assessments, suitable objectives can be identified, and then the training effort can be focused on meeting these objectives. The training efforts can be focused on the Scrum masters, who represent the role of coach (Moe & Dingsøyr, 2008); training efforts can also be focused on teams as a whole to improve shared aspects of SSR. By doing so, it is expected that SSR will be more shared, and will thus become more effective.

Second, a possible implication is that the findings create awareness around the usefulness of high extent sharing of SSR, and around how to share SSR to improve effective SSR. Organizations can profit from our findings by picking out some ideas that are assumed to foster high extent sharedness of SSR. For instance, managers could create awareness in teams about the use of conclusions. The managers can, for instance, demand that agreements are reported, and that each member should announce his vote toward a particular outcome orally. Similar standards can also relate to the distribution of team member contributions during team meetings. It is expected that practice will profit from the insights provided by this study to increase sharedness of SSR.

5. Conclusion

To conclude, this study examined two methods to serve as a measure for the extents of sharedness of SSR in Scrum teams; for effective SSR. The first method had a qualitative approach and seeks to determine the extent to which SSR is shared, based on the extent to which team members use ‘high-level transactive communication’; whereas the second method had a quantitative approach that focused on the extent to which team members established ‘equal team member contributions’. To translate the methods into quantifiable extents of sharedness of SSR, two operationalizations are designed per measurement method. For both methods one operationalization focused on the ‘durations’, and one on the ‘frequencies’. Subsequently, this study examined which method (or combination of methods) is most appropriate to determine the extents of sharedness of SSR.

Findings from the present study indicate that a combination of both measurement methods would generate the most satisfying results regarding the extents of sharedness of SSR. This implies a combination of a method with the qualitative approach, that determines the extent of sharedness of SSR based on the ‘frequencies’ and the ‘durations’ of high-level transactive communication; complemented with a measurement method with a quantitative approach that determines the extents of sharedness of SSR based on the ‘numbers’ of ‘equal team member contributions’.

Additionally, the results revealed that high extents of sharedness of SSR are related to high extent ‘shared conclusion-drawing’. The results did not reveal major differences between ‘explicit’ and ‘implicit conclusion-drawing’. This might mean that the definitions of ‘explicit conclusions’ and ‘implicit conclusions’ should be reconsidered. Nevertheless, this distinction seems adequate to explain variations in the extents of sharedness of SSR.

Finally, both methods did not reveal that variations in the extents of sharedness of SSR are attributable to the ‘type of Scrum meeting’. With the exception of one operationalization of the method with the quantitative approach (based on ‘equal team member contributions’), that determined the extents of sharedness of SSR based on the ‘number’ of team members involved in topic. Possibly, a time-restriction for the meeting, forces employees to more effectively, and efficiently organize team member involvement. Nevertheless, the ‘type of Scrum meeting’ seems inappropriate to explain variations in the extents of sharedness of SSR. Research on SSR can build further on the results to expand knowledge on determining the extent to which SSR is shared among team members. For practice, our finding can help to demonstrate the need to undertake training investments that increase the extents of sharedness of SSR.

References

Baarda. (2010). *This is it! Guidelines for setting up, doing and evaluating quantitative and qualitative research.* Houten, Groningen: Noordhoff B.V.

Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, (1982-), 399-410.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Sage publications.

Cho, J. (2008). Issues and Challenges of agile software development with SCRUM. *Issues in Information Systems*, *9*(2), 188-195.

Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data* (Doctoral dissertation, University of Pittsburgh).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Routledge Academic.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of management*, pp. *23*(3), 239-290.

De Backer, L., Van Keer, H., & Valcke, M. (2015). Socially shared metacognitive regulation during. *Instructional Science*, *43*(3), 323-344. doi:10.1007/s11251-014-9335-4

DiDinato, N. C. (2013). Effective self-and co-regulation in collaborative learning groups: An analysis of how students regulate problem solving of authentic interdisciplinary tasks. *Instructional Science*, 41(1), 25-47.

Dyba, T., & Dingsoyr, T. (2015). Agile Project Management: From Self-Managing Teams to Large-Scale Development. In Software Engineering (ICSE) 2015. 37th IEEE International Conference on (Vol. 2, pp. 945-946).

Dybå, T., Dingsøyr, T., & Moe, N. B. (2014). Agile project management. In G. Ruhe & C. Wohlin (Eds.). *Software Project Management in a Changing World*, (277-300).

Espinosa, J. A., Slaughter , S. A., Kraut, R. E., & Herbsleb, J. D. (2007). Familiarity, complexity, and team performance in geographically distributed software development. *Organization Science, 18*(4), 613-630.

Field, A. P. (2005). Eta and eta squared. *Wiley StatsRef: Statistics Reference Online*.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and Instruction, 22*(6), 401-412.

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264.

Hinkle, D. E., Wiersma , W., & Jurs, S. G. (1988). *Applied Statistics for the Behavioral Sciences* (5 ed.).

Hoegl , M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science, 12*(4), 435-449.

Hossain, E., Babar, M. A., & Paik, H. Y. (2009). Using scrum in global software development: a systematic literature review. *Global Software Engineering*.

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction, 21*(3),

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorrative dentistry & endodontics, 38*(1), *52-54*.

Klonek, F. E., Quera, V., Burba, M., & Kauffeld, S. (2016). Group interactions and time: Using sequential analysis to study group dynamics in project meetings. *Group Dynamics: Theory, Research, and Practice*, pp. 20(3), 209.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association, 47*(260), 583-621.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual review of public health, 23*(1), 151-169.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530.

Mattessich, P. W., & Monsey, B. R. (1994). *Collaboration: what makes it work.* St. Paul, Minnesota: Amherst H. Wilder Foundation.

Meijer, J., Veenman, M. V., & van Hout-Wolters, B. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Evaluation, 12*(3), 209-237.

Moe, N. B., & Dingsøyr, T. (2008). Scrum and team effectiveness: Theory and practice. *Agile Processes in Software Engineering and Extreme Programming, 26*(6), 11-20.

Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Srum project. *Information and Software Technology*, *52*(5), 480-491.

Molenaar, I. (2011). It's all about metacognitive activities: Computerized scaffolding of self-regulated learning. *Ipskamp Drukkers BV*, Unpublished doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands.

Muller, A. (2017). Socially shared regulation in Scrum teams: A multiple case study into the regulation processes in Scrum team meetings and variations both within and across teams. (Master's thesis, University of Twente).

Raes, E., Boon, A., Kyndt, E., & Dochy, F. (2015). Measuring team learning behaviours through observing verbal team interaction. *Journal of Workplace Learning*, *27*(7), 476-500.

Rising, L., & Janoff, N. S. (2000). The Scrum software development process for small teams. *IEEE software*, *17*(4), 26-32.

Rogat, T. K., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction*, *29*(4), 375-415.

Saarni, C. (1982). Social and affective functions of nonverbal behavior: Developmental concerns. In *Development of nonverbal behavior in children* (pp. 123-147). Springer, New York, NY.

Schoor, C., Narciss, S., & Körndle, H. (2015). Regulation During Cooperative and Collaborative Learning: A Theory-Based Review of Terms and Concepts. *Educational Psychologist*, *50*(2), 97-119.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, *6*(1), 35-39.

Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*.

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? 361-384.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review*, *3*(2), 130-154.

Van der Haar, S. , Wijenberg, B. , Van den Bossche, P. and Segers, M. (2013). Team learning behavior: A study in the setting of command and control teams. Paper presented at the 15the Biennial Conference Earli 2013, Munich, Germany.

Vaura, M., Iiskala, T., Kajamies, A., & Kinnunen, R. (2003). Shared-regulation and motivation of collaborating peers: A case analysis. *Psychologia*, *46*(1), 19-37.

Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, *19*(2), 128-143.

Volet, S., Vauras, M., & Salonen, P. (2009). Self-and social regulation in learning contexts: An integrative perspective. *Educational psychologist*, 44(4), 215-226.

Wijga, M., & Endedijk, M. D. (2016). Social Regulation at the Workplace: The Development of an Analysis Tool to Understand Regulation of Learning in Teams. Paper presented at the International Conference: 8th EARLI SIG 14 Learning and Professional Development, Regensburg. http://doc.utwente.nl/102694/

Wiersma, S. G., Jurs, S. G., & Hinkle, D. E. (2003). *Applied Statistics for the Behavioral Sciences* (5 ed.). Boston: Houghton Mifflin.

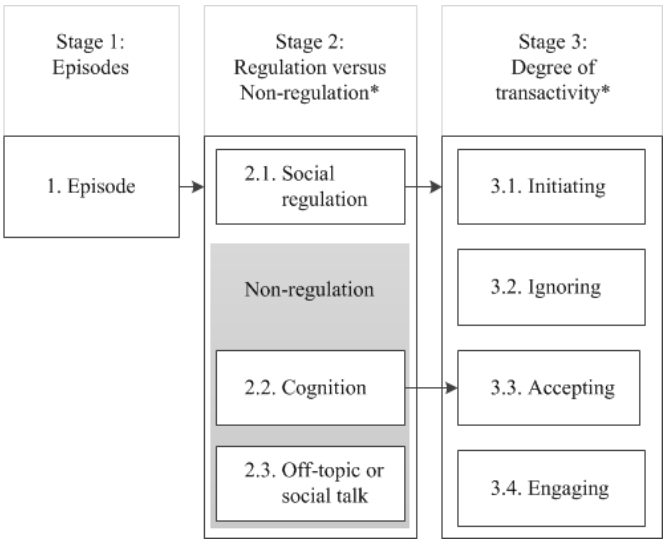
Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329.

Appendix A Coding Scheme 'High-level Transactive Communication'

The coding process is illustrated by Figure 2. Table 9 present descriptions on the coding categories with examples.

Figure A1

Schematic Representation of the Coding Scheme



Note. The extents of sharedness of SSR are coded in three Stages. *Coded on the utterance level. Stage 3 and 4 are removed from the original scheme because these are not relevant four our current study.

Stage 1: Episodes. The episodes are segmented (see Stage 1 in Table 9). An episode contains a sequence of utterances focused around one particular topic or discussion. The start of an episode is denoted by the "Initiating" code (see number 3.1 in Table 9), that indicated a change of topic. The "Initiating" code is assigned to each first utterance of the 'new' topic.

Stage 2: Regulation versus Non-regulation. The utterances *regulation utterances* are distinguished from *non-regulation utterances* (see Stage 2 in Table 9). Regulation utterances are directed towards 'planning', 'monitoring' and 'evaluating' (see number 2.1 in Table 9). Whereas non-regulatory utterances are focused on their affordances to discuss content-related issues ('cognition'), off-topic issues, or the content cannot be clearly understood. These utterances are split in two categories: "Cognition" (see number 2.2 in Table 9) and "Off-topic or social talk" (see number 2.3 in Table 9). The 'social regulation' utterances are further specified in Stage 3. Concerning the *non-regulation utterances*, the content of the "Off-topic or social talk" category is not further coded.

Stage 3: Degree of transactivity. To determine the extents of sharedness of SSR based on the degree of transactivity (see Stage 3 in Figure 2). The original codes are used. The 'engaging' code is only regarded as high extent sharing of SSR.

Table A1

Codes Type of Transact with Definition and Example

Stage	Code	Definition	Example
Stage 1: Episodes.			
	Episode	A sequence of utterances about the same topic	Episode about which tasks to perform first this sprint
Stage 2: Regulation versus Non-regulation.			
2.1	Social regulation	Intentional and goal directed group efforts to regulate its conceptual understanding and task work. Collectively shared regulatory processes orchestrated in the service of shared outcome.	TM a: Let's start this meeting with…"
2.2	Cognition	Utterances about the content of the task and the elaboration of this content.	TM a: I find it hard to work with the new software.
2.3	Off-topic	When communication is too hard to understand or the sound is unclear.	-
2.3	Social talk	Talk not aimed at regulating the project and the team processes.	TM a: I know a joke to, do you know what I mean?
Stage 3: Degree of transactivity.			
3.1	Initiating	Initiating regulation activity. This can coincide with the ignoring code when the initiating activity is at the same time an ignoring activity of the previous activity.	TM a: *"Are there any questions about this?"* TM b/c/d: (the others remain silent) TM a: *"Ok, good. Then we go further with planning the next Sprint."* (TM a introduces a **new topic**)
3.2	Ignoring	When the group members do not relate to nor engage in another group member's regulation activity.	
3.3	Engaging	When group members relate or engage in each other's regulation activities. Responding by further specifying or clarifying the previous regulation activity or further develop the previously initiated idea.	TM a: "Then we are all present if it is good." TM b: "Yes, that should be the case."
3.4	Accepting	When the group members engage in a regulation activity with a cognitive activity. E.g., confirming, repeating, performing an activity on the spot (answering a question with a non-regulation activity).	TM a: *"I think, last night you have integrated about everything."* TM b: *"Yes."* (TM b **confirms** TM a's contribution, with a short confirmation)

Note. Stage 3 and 4 are removed from the original scheme because these are not relevant four our current study.

Appendix B Raw Extents of Sharedness of SSR

Table B1

Raw Extents of Sharedness of SSR per Operationalization

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
1	DS	A	.00	.00	.18	.40
2	DS	A	.00	.00	.96	.20
3	DS	A	.00	.00	.14	.40
4	DS	A	.00	.00	.08	.20
5	R	A	.28	.60	.00	.60
6	R	A	.00	.00	.00	.20
7	R	A	.66	.50	.31	.40
8	R	A	.00	.00	.00	.20
9	R	A	.78	.73	.90	.80
10	R	A	.80	.50	.00	.60
11	R	A	.93	.90	.22	.80
12	R	A	.65	.50	.00	.40
13	R	A	.00	.00	.64	.20
14	R	A	.00	.00	.00	.20
15	R	A	.00	.00	.07	.20
16	R	A	.92	.94	.14	.80
17	R	A	.76	.60	.12	.80
18	R	A	.00	.00	.54	.40
19	R	A	.00	.00	.00	.20
20	R	A	.00	.00	.00	.40
21	R	A	.00	.00	.00	.20
22	R	A	.28	.60	.24	.60
23	R	A	.00	.00	.13	.20
24	R	A	.00	.00	.00	.20
25	R	A	.91	.96	.26	1.00
26	R	A	.60	.87	.28	.40
27	R	A	.71	.76	.14	.80
28	R	A	.73	.87	.13	.60

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
29	R	A	.67	.56	.86	.40
30	R	A	.75	.94	.02	.60
31	R	A	.60	.73	.23	.60
32	R	A	.00	.00	.00	.20
33	R	A	.00	.00	.00	.25
34	R	A	.50	.35	.18	.50
35	R	A	.83	.86	.11	.75
36	R	A	.83	.93	.32	.50
37	R	A	.83	.95	.47	.50
38	R	A	.89	.94	.30	1.00
39	R	A	.50	.62	.34	.50
40	R	A	.92	.84	.19	.75
41	R	A	.85	.94	.02	1.00
42	R	A	.33	.18	.66	.50
43	DS	A	.50	.80	.25	.50
44	DS	A	.79	.86	.05	1.00
45	DS	A	.50	.75	.34	.40
46	DS	A	.71	.57	.12	.60
47	DS	A	.75	.73	.57	1.00
48	DS	A	.80	.90	.28	.50
49	DS	A	.71	.70	.77	.50
50	DS	A	.00	.00	.66	.50
51	DS	A	.73	.80	.01	1.00
52	DS	A	.39	.50	.26	.50
53	DS	A	.83	.67	.37	.50
54	DS	A	.27	.50	.04	.50
55	DS	A	.51	.33	.92	.50
56	DS	A	.82	.69	.06	1.00
57	DS	A	.00	.00	.39	.50
58	DS	A	.71	.69	.46	.75
59	DS	A	.00	.00	.04	1.00
60	DS	A	.75	.60	.42	1.00

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
61	DS	A	.00	.00	.03	1.00
62	DS	A	.00	.00	.00	.50
63	DS	A	.00	.00	.00	.50
64	DS	A	.00	.00	.00	.50
65	DS	A	.61	.50	.65	1.00
66	DS	A	.69	.63	.55	1.00
67	DS	A	.76	.50	.24	.67
68	DS	A	.65	.67	.10	1.00
69	DS	A	.57	.60	.39	.67
70	DS	A	.81	.67	.27	1.00
71	DS	A	.12	.50	.00	.33
72	DS	A	.73	.50	.00	.33
73	DS	A	.00	.00	.00	.33
74	DS	A	.92	.83	.68	.67
75	DS	A	.81	.78	.11	.50
76	DS	A	.72	.86	.21	.67
77	DS	A	.91	.82	.04	1.00
78	DS	A	.56	.67	.05	.33
79	DS	A	.26	.25	.07	.50
80	DS	A	.00	.00	.00	.25
81	DS	A	.00	.00	.11	.50
82	DS	A	.00	.00	.09	.50
83	DS	A	.76	.61	.16	1.00
84	DS	A	.76	.80	.92	.50
85	DS	A	.65	.60	.28	.75
86	DS	A	.88	.70	.13	1.00
87	DS	A	.92	.83	.14	.75
88	DS	A	.83	.64	.13	.75
89	DS	B	.00	.00	.01	.50
90	DS	B	.68	.56	.03	1.00
91	DS	B	.95	.75	.16	1.00
92	DS	B	.80	.67	.53	.50

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
93	DS	B	.25	.60	.16	.50
94	DS	B	.64	.56	.08	.75
95	R	B	.00	.00	.00	.14
96	R	B	.37	.60	.13	.43
97	R	B	.72	.67	.34	.43
98	R	B	.87	.80	.17	.29
99	R	B	.94	.77	.05	.71
100	R	B	.44	.57	.35	.43
101	R	B	.80	.89	.56	.43
102	R	B	.00	.00	.14	.29
103	R	B	.78	.73	.14	.71
104	R	B	.95	.75	.25	.29
105	R	B	.75	.64	.08	.57
106	R	B	.00	.00	.00	.14
107	R	B	.83	.73	.01	.86
108	R	B	.97	.88	.02	1.00
109	R	B	.94	.88	.34	.71
110	R	B	.00	.00	.00	.14
111	R	B	.54	.50	.35	.43
112	R	B	.00	.00	.00	.14
113	R	B	.85	.75	.09	.43
114	R	B	.61	.63	.02	.57
115	R	B	.90	.84	.03	1.00
116	R	B	.76	.87	.22	.86
117	R	B	.92	.85	.37	.57
118	R	B	.93	.86	.22	.71
119	R	B	.00	.00	.00	.14
120	DS	B	.11	.33	.19	.50
121	DS	B	.72	.67	.56	.75
122	DS	B	.55	.50	.42	.75
123	DS	B	.66	.57	.78	.50
124	DS	B	.47	.50	.24	.50

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
125	DS	B	.56	.50	.20	1.00
126	DS	B	.23	.50	.09	.50
127	DS	B	.66	.83	.17	.75
128	DS	B	.60	.50	.18	.50
129	DS	B	.00	.00	.11	.50
130	DS	B	.00	.00	.48	.50
131	DS	B	.64	.57	.31	1.00
132	DS	B	.00	.00	.73	.50
133	DS	B	.91	.86	.06	.75
134	DS	B	.90	.60	.11	.75
135	DS	B	.34	.75	.04	.75
136	DS	B	.45	.40	.95	.50
137	DS	B	.54	.67	.16	.50
138	DS	B	.70	.57	.60	.50
139	DS	B	.58	.50	.11	.75
140	DS	B	.67	.33	.03	.67
141	DS	B	.08	.33	.03	.67
142	DS	B	.71	.60	.65	.67
143	DS	B	.49	.33	.97	.67
144	DS	B	.00	.00	.00	.33
145	DS	B	.85	.60	.23	.67
146	DS	B	.82	.57	.59	.67
147	DS	B	.93	.50	.07	.67
148	DS	B	.63	.40	.70	1.00
149	DS	B	.00	.00	.00	.33
150	DS	B	.98	.87	.27	1.00
151	DS	B	.61	.60	.21	1.00
152	DS	B	.50	.67	.08	1.00
153	DS	B	.15	.33	.00	1.00
154	DS	B	.80	.67	.46	1.00
155	DS	B	.00	.00	.28	1.00
156	DS	B	.85	.67	.43	1.00

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
157	DS	B	.43	.60	.11	1.00
158	R	C	.00	.00	.00	.33
159	R	C	.00	.00	.01	.67
160	R	C	1.00	.90	.04	1.00
161	R	C	.94	.82	.02	1.00
162	R	C	.77	.76	.41	1.00
163	R	C	.69	.50	.46	.67
164	R	C	.56	.69	.68	.67
165	R	C	.92	.67	.05	.67
166	R	C	.67	.50	.00	.33
167	R	C	.66	.50	.00	.33
168	R	C	.00	.00	.00	.33
169	R	C	.00	.00	.70	.67
170	R	C	.65	.75	.34	1.00
171	R	C	.32	.67	.11	.40
172	R	C	.24	.50	.32	.40
173	R	C	.00	.00	.27	.40
174	R	C	.97	.77	.01	1.00
175	R	C	.75	.69	.02	1.00
176	R	C	.95	.71	.05	1.00
177	R	C	.93	.89	.02	.60
178	R	C	.59	.50	.69	.40
179	R	C	.00	.00	.00	.20
180	DS	C	.81	.50	.04	.75
181	DS	C	.00	.00	.00	.25
182	DS	C	.19	.50	.20	.50
183	DS	C	.94	.87	.03	1.00
184	DS	C	.82	.50	.22	.50
185	DS	C	.92	.60	.17	.50
186	DS	C	.52	.33	.17	.75
187	R	C	.78	.67	.53	1.00
188	R	C	.97	.85	.16	1.00

Episode	Meeting Type	Team	Rd HLTC	Rf HLTC	Rd CTM	Rn CTM
189	R	C	.99	.86	.06	1.00
190	R	C	.56	.60	.04	1.00
191	DS	C	.88	.57	.01	1.00
192	DS	C	.71	.78	.01	1.00
193	DS	C	.36	.67	.18	.67
194	DS	C	.32	.60	.50	.67
195	DS	C	.50	.50	.00	.33
196	DS	C	.00	.00	.00	.33
197	DS	C	.85	.85	.09	1.00
198	DS	C	.00	.00	.04	.67
199	DS	C	.85	.60	.37	.67
200	DS	C	.93	.50	.06	.67
201	DS	C	.00	.00	.00	.33
202	DS	C	.19	.50	.01	1.00
203	DS	C	.68	.71	.52	.67
204	DS	C	.41	.80	.12	.67
205	DS	C	.96	.50	.04	.67
206	DS	C	.00	.00	.44	.67
207	DS	C	.91	.76	.02	.80
208	DS	C	.79	.83	.04	.60
209	DS	C	.73	.67	.46	.40
210	DS	C	.80	.84	.48	1.00
211	DS	C	.71	.75	.17	.80
212	DS	C	.56	.75	.03	1.00
213	DS	C	.00	.00	.00	.20

Note. DS = daily Scrum; R = Retrospective; Rd HLTC = Relative duration High-Level Transactive Communication; Rf HLTC = Relative frequency High-Level Transactive Communication; Rd CTM = Relative duration Contributing Team Members; Rn CTM = Relative number Contributing Team Members.