Empir Software Eng
DOI 10.1007/s10664-011-9179-3

Computer-mediated communication to support distributed requirements elicitations and negotiations tasks

Fabio Calefato · Daniela Damian · Filippo Lanubile

Editor: Daniel M. Berry

Abstract Requirements engineering is one of the most communication-intensive activities in software development, greatly affected by project stakeholder geographical distribution. Despite advances in collaboration technologies, global software teams continue to experience significant challenges in the elicitation and negotiation of requirements. Deciding which communication technologies to deploy to achieve effective communication in distributed requirements engineering activities is not a trivial task. Is face-to-face or text-based communication more appropriate for requirements elicitations and negotiations? In teams that do not have access to face-to-face communication, is text-based communication more useful in requirements elicitations than in requirements negotiations? Here, we report an empirical study that analyzes the effectiveness of synchronous computer-mediated communication in requirements elicitations and negotiations. Our investigation is guided by a theoretical framework that we developed from theories of computer-mediated communication, common ground, and media selection for group tasks; a framework that considers the effectiveness of a communication medium in relation to the information richness needs of requirements elicitation and negotiation tasks. Our findings bring forward empirical evidence about the perceived as well as objective fit between synchronous communication technology and requirements tasks. First, face-to-face is not always the most preferred medium for requirements tasks, and we reveal a number of conditions in which, in contrast to common belief, text-based communication is preferred for requirements communication. Second, we find that in evaluating outcomes of requirements elicitations and negotiations objectively, group performance is not affected by the communication medium. Third, when groups interact only via text-based communication, common ground in requirements

F. Calefato · F. Lanubile (✉)
Dipartimento di Informatica, University of Bari, Bari, Italy
e-mail: lanubile@di.uniba.it

F. Calefato
e-mail: calefato@di.uniba.it

D. Damian
Department of Computer Science, University of Victoria, Victoria, BC, Canada
e-mail: danielad@cs.uvic.ca

Published online: 01 October 2011

negotiations takes longer to achieve than in requirements elicitations, indicating that distributed requirements elicitation is the task where computer-mediated communication tools have most opportunity for successful application.

Keywords Global software development · Requirements elicitations · Requirements negotiations · Computer-mediated communication · Synchronous media selection · Task-technology fit · Empirical study

1 Introduction

Effective communication is crucial to system design (Curtis et al. 1988). Especially at the requirements stage, system design is a social and communication-intensive activity that relies on an effective collaboration of stakeholders with diverse professional and cultural backgrounds (Cheng and Atlee 2007; Curtis et al. 1988; Nuseibeh and Easterbrook 2000). Whether engineered or naturally emerging and agreed upon during a negotiation process, requirements demand increased communication during elicitation and negotiation. Effective communication is vital during these activities to overcome the semantic gap between users and designers, as well as to reconcile the aspects of the design process affected by human and organizational factors (Coughlan and Macredie 2002; Macaulay 1996). Communication problems in requirements engineering manifest in problems in articulation (expressing requirements), in misunderstandings, and in conflicts (Sutton 2000; Walz et al. 1993), and negatively affect the shared understanding of users and designers, presenting obstacles to the success of a system.

Practical considerations make the study of computer-mediated communication for requirements tasks appealing. Computer-mediation has the potential to overcome problems of group dynamics in large groups (Gottesdiener 2002); as well, software teams increasingly develop software in predominantly distributed settings and rely on computer-mediated collaborative tools to mediate their design activities. Geographical, organizational, and cultural distance brings additional challenges to effective communication and results in misunderstandings (Damian and Zowghi 2003; Damian 2007), the loss of opportunities for rich interaction, and a reduction in frequency of both formal and informal communication (Carmel and Agarwal 2001; Herbsleb et al. 2001).

Accordingly, it would be useful to further our understanding of the effectiveness of synchronous communication media to support requirements activities in software teams, especially those working in geographically distributed settings. This paper describes research that examines synchronous text-based communication in comparison to face-to-face communication, regarded as the richest and most preferred mode of communication. Previous work in the area of computer-mediated support for distributed requirements engineering has evaluated factors affecting the quality of requirement elicitations conducted through audioconferencing (Lloyd et al. 2002) and has compared group performance in videoconferencing vs. face-to-face meetings (Damian et al. 2000). Additionally, literature reports studies of human facilitation of videoconferencing-supported distributed requirements meetings (Damian et al. 2003); and evaluations of the feasibility of synchronous or asynchronous requirements negotiations of distributed groups (Campbell and Van de Walle 2003; Damian et al. 2008). Synchronous text-based communication appears to be understudied in distributed software teams, despite its relative low cost compared to the cost of audio or videoconferencing systems and its wider availability in recent collaborative tools (e.g., Skype, Google Talk).

In Section 2, we discuss the properties of collaboration in requirements engineering and its computer-mediated communication and motivate our focus on the tasks of requirements elicitation and negotiation. In Section 3, we review the most prominent theories of communication media fit to group tasks. Then, in Section 4, we develop a theoretical framework that draws upon existing theories of computer-mediated communication and media selection. With this framework, we analyze the task/technology fit for distributed activities of requirements elicitations and negotiations and develop a number of hypotheses about the effectiveness of synchronous, distributed text-based requirements elicitations and negotiations in comparison with their face-to-face counterparts. We test these hypotheses in an empirical investigation of which, in Section 5, we describe the settings and instrumentation. Our findings are discussed in Section 6 and we conclude with implications of our research in Section 7.

2 Requirements Engineering and Computer-mediated Communication

Requirements Engineering (RE) is an important domain for studying computer-mediation in distributed software teams. Collaboration in RE demands a complex interplay between cognitive and behavioral processes that underlie idea generation, decision-making, and requirements conflict resolution (Nuseibeh and Easterbrook 2000). Techniques that foster effective communication and agreement on requirements include group elicitation techniques such as group brainstorming, Joint Application Design/Development (JAD) sessions (Andrews 1991), and focus groups (Macaulay 1996); or requirements creativity workshops (Maiden et al. 2004). The techniques foster stakeholder mutual understanding, decision-making, agreement, and buy-in, while exploiting team dynamics to elicit and negotiate a rich understanding of user needs (Nuseibeh and Easterbrook 2000; Gottesdiener 2002). However, team dynamics can also be the source of social pressure and "groupthink" that affect decision-making and negotiations (Gottesdiener 2002; Macaulay 1996; Sommerville and Sawyer 1997). Global teams suffer from additional challenges, including the inability to participate due to geographical or temporal distance, or language differences, which lower the ability to reach shared understanding (Damian 2007). In distributed requirements elicitations and negotiations (referred to as *requirements workshops* henceforth), collaborative technologies for synchronous and asynchronous communication show potential to mitigate some of these problems and to increase the effectiveness of computer-mediated distributed requirements workshops.

Literature (e.g., Gottesdiener 2002) has documented some of the expected consequences of collaborative technologies for RE tasks. Their benefits include increased speed and immediacy of documentation, focused discussions, and greater participation of larger groups. However, such mediation may also include loss of human contact and non-verbal cues, longer time in building common ground, and difficulties in reaching closure. Empirical studies show that these benefits/minuses are realized differently depending on the specific RE activity performed during requirements workshops such as elicitations (Lloyd et al. 2002) or negotiations (Campbell and Van de Walle 2003; Damian et al. 2000; Damian et al. 2003; Damian et al. 2008). Elicitations and requirements negotiations are two communication-intensive activities that differ in their information needs for effective performance and in their demands for media richness. Whereas elicitations are intended to be predominantly intellectual, idea generation, and knowledge acquisition tasks, requirements negotiation tasks relate to conflict resolution, knowledge, and to solution negotiation (Nuseibeh and Easterbrook 2000; Sommerville and Sawyer 1997). Early requirements

workshops in a project may focus on gathering information from the stakeholders (elicitation tasks), whereas later workshops may follow requirements analysis and modeling tasks and focus on stakeholder agreement on requirements and solution (negotiation tasks). Deciding which communication technology to employ to achieve the best task/technology fit in terms of the effective performance of these tasks is not trivial. How does one decide which communication media (e.g., face-to-face meetings, audioconferencing, videoconferencing, or simply text-based communication) is best to mediate requirements elicitation tasks? What about mediating requirements negotiation tasks? Can we anticipate the performance of groups in requirements workshops based on the affordances inherent in a particular communication medium *and* the requirements task's communication needs?

Guided by a theoretical framework described in Section 4, our empirical investigation analyzes computer-support for requirements elicitations and negotiations. The significant, but controversial, body of knowledge in the field of computer-mediated communication and media selection theories indicates that no theory can be accepted or considered valid *tout court* and that only an analysis of the appropriateness of the fit between task characteristics and media characteristics will get the best out of media use (Zigurs and Buckland 1998). We review these theories in the next section. The framework we develop draws on this work by analyzing the specific characteristics of requirements elicitation and negotiation tasks and synthesizing what could be the best technology fit for these tasks. We capture this fit in the three hypotheses of our empirical study described in the remaining sections of the paper.

3 Theories of Computer-mediated Communication and Task/technology Fit

Media *richness* can be intuitively defined as the ability of communication media to convey a large amount of information in different forms. Face-to-face (F2F) interaction is considered the richest, followed by videoconferencing, audioconferencing, and written text as the leanest. Many computer-mediated communication (CMC) theories (e.g., Short et al. 1976; Daft and Lengel 1986) have provided various different definitions of media richness but, despite differences, the resulting ranking of media richness does not change from that presented above. However, a point the CMC theories have in common is the inadequacy of lean media for complex, collaborative tasks; as task complexity increases, richer media are suggested as appropriate (Short et al. 1976; Daft and Lengel 1986; Clark and Brennan 1991).

The concept of media selection (prescribing the best media for executing tasks) is the basis of theories of computer-mediated communication (CMC). Theories such as *Social Presence* (Short et al. 1976) and *Media Richness* (Daft and Lengel 1986) hypothesize that group effectiveness decreases when media other than F2F are used to accomplish *equivocal* tasks that require the exchange of interpersonal cues. *Equivocality,* defined as the existence of multiple and conflicting interpretations of a situation (Daft and Lengel 1986), is symptomatic of ambiguity, confusion, disagreement, and lack of understanding. Because exploring different viewpoints and reaching closure reduces equivocality, the exchange of interpersonal cues is helpful in such contexts. These theories also report the inadequacy of text-based communication in comparison with rich media (i.e., F2F and video) in supporting performance of equivocal tasks. Lean media, such as e-mail and instant messaging, lack the ability to convey nonverbal cues that contribute to *social presence* (e.g., gaze, tone of voice, facial expressions); social presence, in turn, fosters individual exchange of interpersonal cues and participation in discussion. Social presence is a strong

indicator of satisfaction; that is, the higher the level of social presence conveyed by media, the higher the level of satisfaction perceived (Gunawardena and Zittle 1997). At the same time, however, Social Presence and Media Richness theories hypothesize that lean media can increase effectiveness of groups in *task-focused* collaboration when tasks are *uncertain*, rather than equivocal, and when the exchange of interpersonal cues is less vital. *Uncertainty* is defined as the discrepancy between the amount of information required to perform a task and the amount already possessed (Daft and Lengel 1986). Because obtaining additional information reduces uncertainty, text-based media that enhance an individuals' ability to process additional information is helpful.

Similar recommendations of task-technology fit are provided by *Common Ground* theory (Clark and Brennan 1991). The *common ground*, or shared understanding, is the knowledge—and the awareness of it—that participants have in common when communicating. Not necessarily based on previous knowledge, it is possible to establish common ground dynamically through an interactive process called grounding, through which participants exchange evidence about what they do or do not understand over the course of a conversation and interpret cues obtained during social interaction. Early experiments suggest that people who have little common ground profit from rich, audio- and video-based communication channels in establishing common ground necessary in task completion; the greater the task equivocality, the greater the need for effective achievement of common ground in converging to a shared view (Veinott et al. 1999). Conversely, people who have an extensive preexisting common ground can communicate effectively over lean, text-based media, such as e-mail (Olson et al. 1995; Olson and Olson 2000). However, recently Birnholtz et al. (2005) reported the existence of collaboration settings, characterized by reduced information loads, where synchronous, text-based communication was adequate to achieve common ground among conversational participants unknown to each other, and with a low level of initial common ground.

Although useful, early CMC theories have strong face validity, but empirical evidence is rather equivocal. Usually supported when tested on traditional media, such as F2F communication and telephone, reports indicate inconsistent empirical findings when email and video media are employed (Dennis and Valacich 1998). In addition, research on Group Support Systems (GSS; for an exhaustive compendium, see Fjermestad and Hiltz 1998) has consistently reported that distributed groups interacting via text-chat outperform collocated groups in idea-generation tasks, but are in turn outperformed by collocated groups in problem-solving tasks (Murthy and Kerr 2000). Nevertheless, performance of distributed groups, collaborating over a computer-mediated medium, was often perceived as lower than that of collocated groups, even when it was objectively measured to be equal if not superior. This is because individual perception of performance is biased by the higher level of *comfort* perceived when producing and receiving messages over rich audio/video channels.

These inconsistencies have encouraged a reconsideration of the descriptive and predictive general validity of early CMC theories. Recent theories such as *Media Synchronicity* (Dennis et al. 2008) and the *Cognitive-Based View* (Robert and Dennis 2005) posit that the effectiveness of computer-mediated communication depends on factors other than media richness. The Cognitive-Based View theory considers communication as a *cognitive process*. Apart from a sender's comfort with the communication medium, a receivers' *commitment* (i.e., motivation, attention) and ability to process the message properly are equally important. The Cognitive-Based View theory argues that rich media will ensure high commitment for exchanging small amounts of information. Lean media on the other hand, while causing a likely decrease in *motivation* and *attention*, provide

increased support for *reprocessability*, that is, the ability to (re)process at will large amounts of information during longer periods of time (Robert and Dennis 2005). Similarly, Media Synchronicity theory distinguishes between two fundamental communication processes, namely, conveyance and convergence. *Conveyance* is the exchange of information followed by reflection on its meaning. Conveyance involves *divergent thinking* (i.e., creativity, generating ideas), which means that not all participants focus on the same information at the same time, nor must they agree on its meaning. Instead, c*onvergence* is the development of shared meaning for information. By definition, it involves *convergent thinking* (i.e., participants explore and understand one another's views), in that participants strive to agree both on the meaning of information and that they have reached closure. The support for conveyance and convergence varies with the degree of *synchronicity* of the medium. A medium's degree of synchronicity is the extent to which it supports *immediacy of feedback* (the ability of a medium to have rapid bidirectional communication) and *input parallelism* (the ability of a medium to allow more simultaneous conversations at a time), in an inversely proportional relationship. For example, F2F conversation is highly synchronous because it grants high immediacy of feedback, with no support for parallel input (i.e., only one conversation at a time). In contrast, instant messaging is less synchronous than F2F conversation, because it supports lower immediacy of feedback, but ensures higher input parallelism. Media Synchronicity theory suggests that when extra information (i.e., conveyance) is needed, media that are less synchronous are a better fit due to the support of input parallelism; whereas highly synchronous media can better support convergence because of the higher degree of immediacy of feedback.

The concept of alignment between task and media characteristics is also the focus of the media selection theories of *Time-Interaction-Performance* (McGrath 1991) and *Task/ Technology Fit* (Zigurs and Buckland 1998). These theories evaluate the appropriateness of task-medium matches, considering tasks not as somewhat atomic activities, as do all media richness theories, but, rather, as complex sets of sub-tasks and sub-processes, each having different characteristics. Both theoretical frameworks argue that rich media do not always provide the best-fitting combination regardless of the task type because task-media fits are appropriate only when the level of information richness of a medium is adequate to the complexity of the task. In other words, these theories posit the existence of two types of poor-fit combinations. In the first, tasks require more information richness than selected media can deliver and groups may suffer from problems of effectiveness and quality, forcing individuals to exchange further compensative information. In the second, media provide more information richness than tasks require and groups may suffer from problems of efficiency because media convey not only facts, but also non-essential communication (e.g., interpersonal and affective messages), which brings distraction. In particular, Task/ Technology Fit theory defines opportune fit profiles to align task characteristics to several dimensions of technology. Because we are mainly interested in the intrinsic property of communication media, here we consider only the *group communication support* dimension. Lean media, such as text-based conferencing, have a high group communication support level, as they typically grant at the same time input feedback, simultaneous input, and group display. These features are not enabled when using rich media, such as F2F or videoconferencing, which, therefore, grant low group communication support. Primarily due to simultaneous input, the theory prescribes the use of media with high group communication support for tasks that focus on generating ideas and reducing the uncertainty in information associated with the tasks themselves. Conversely, to avoid the risk of communication overload, media with low group communication support are preferred for tasks that focus on reaching closure.

4 Theoretical Framework and Study Hypotheses

With these theories of computer-mediated communication and task/technology fit, we develop a framework of analysis of computer support for requirements elicitations and negotiations (for simplicity, henceforth referred to as requirements tasks) to inform the hypotheses studied in our empirical study. Illustrated in Fig. 1, our framework contrasts the properties of lean media (such as text-based communication, TXT henceforth), with that of rich media (such as face-to-face communication, F2F henceforth). This contrast rests on the CMC theories and properties relevant to the analysis of task/technology fit as shown in the center of the figure. Based on these theories, two levels emerge as important in the analysis of the CMC technology fit to the performance of requirements tasks. These two levels distinguish between *appreciation,* a rather *subjective* task/technology fit, and *effectiveness,* a more *objective* task/technology fit. We also consider tasks along two dimensions, according to McGrath's classification of tasks (McGrath 1984) (see Zigurs and Buckland 1998 for a list of historical task classification frameworks): the degree to which processes involve *cooperative* (i.e., low member interdependence) versus *conflictual* (i.e., high member interdependence) activities, and the degree to which the processes involve *conceptual* versus *behavioral* activities.

The first level of analysis is informed by the early media richness theories, such as Social Presence (Short et al. 1976) and Media Richness (Daft and Lengel 1986), which distinguish between lean and rich media without special regard to the task at hand. The rich media are *highly synchronous* (Dennis et al. 2008); convey, as compared to lean media, a

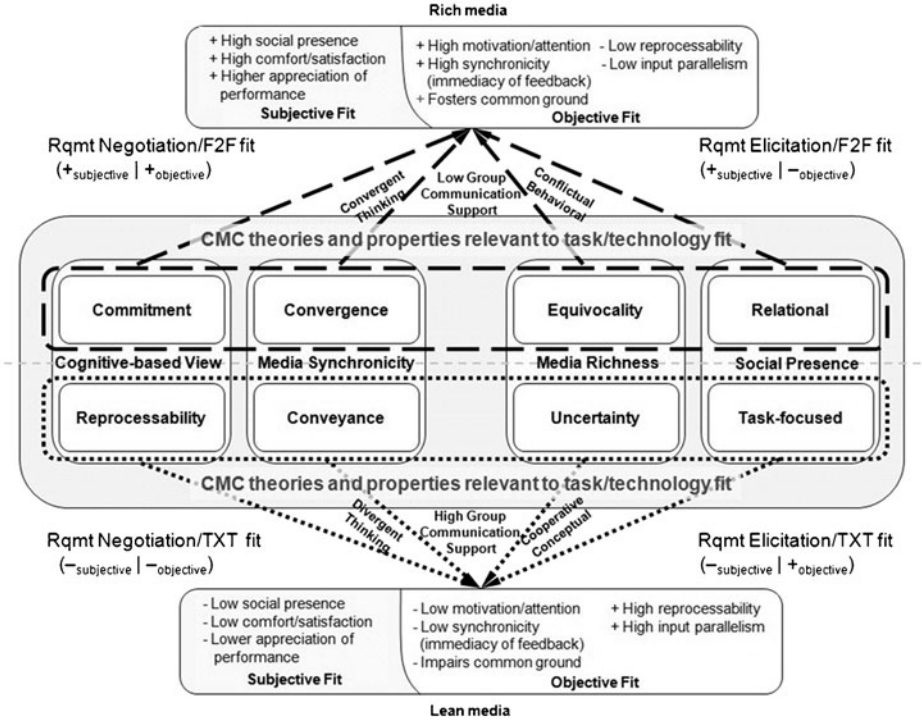


Fig. 1 Task/technology fits as hypothesized by theories on CMC for requirements elicitations and negotiations

higher sense of *social presence* (Short et al. 1976); ensure a higher level of *commitment* (i.e., attention and motivation) (Robert and Dennis 2005); and facilitate mutual understanding by fostering the establishment of *common ground* (Clark and Brennan 1991). These characteristics lead to a higher level of comfort and satisfaction during an interaction mediated by rich media. Conversely, lean media are at the other end of this spectrum. Whatever the task, the *appreciation of performance* when collaborating over lean media is always biased by the lower *comfort* perceived in comparison with rich media. Consequently, from this perspective, the performance of requirements tasks using computer-mediated communication could be perceived as lower than when using rich media such as face-to-face meetings. Our first hypothesis indicates that both F2F requirements elicitations and negotiations are good *subjective* fits ($+_{\text{subjective}}$), and both TXT requirements elicitations and negotiations are poor *subjective* fits ($-_{\text{subjective}}$):

H1 *F2F requirements elicitations (F2F negotiations) are better appreciated than TXT requirements elicitations (TXT negotiations).*

The second level of analysis depends on more recent theories of communication, theories that allow us to consider task characteristics in an analysis of the effectiveness of computer-mediated performance of requirements tasks. Certain properties of the communication media and their interaction with those of the task at hand allow us to speak about the more *objective effectiveness* (rather than perceived effectiveness) of the specific computer mediated requirements task, and to classify the task/technology fits in requirements elicitations and negotiations as poor ($-_{\text{objective}}$) or good ($+_{\text{objective}}$).

For the purpose of this analysis, we consider the requirements elicitation task as a type of *task-focused*, *conceptual* activity of idea generation involving *cooperation* rather than conflict or competition, and which requires limited member interdependence (Macaulay 1996; Nuseibeh and Easterbrook 2000; Sommerville and Sawyer 1997). From a cognitive perspective, the process of generating new ideas primarily involves *divergent thinking* (Dennis et al. 2008). Building then on the Social Presence theory (Short et al. 1976), one would assume that the divergent thinking process requires little need for communicating social information, which in turn tends to make participants more susceptible to pressure of social consensus and domination, and to take time away from task-oriented interaction.

In fact, task/technology fit theories consistently argue that using media that provide non-essential communication to task execution generates poor fits due to efficiency problems (McGrath 1991; Zigurs and Buckland 1998). As well, as stated by Media Richness theory (Daft and Lengel 1986), it is possible to reduce the *uncertainty* inherent in a generative task by conveying additional information. According to Media Synchronicity theory (Dennis et al. 2008), the *conveyance* of extra information necessary to reduce the *uncertainty* inherent in a generative task is better supported by lean media, which are high in parallelism and which foster idea generation by allowing multiple individuals to contribute information simultaneously. In fact, compared to rich media, lean media ensure *high group communication support* (Zigurs and Buckland 1998), which allows multiple input (high parallelism), and *high reprocessability* (Robert and Dennis 2005), allowing individuals to deliberate at will on the information exchanged. Conversely, rich media offer low group communication support (low parallelism) (Zigurs and Buckland 1998) and low reprocessability (Robert and Dennis 2005), in addition to the potential for information overload, due to the multiple channels available at once to convey information. Thus, the framework indicates requirements elicitations performed on lean media as a good objective task/

technology fit ($+_{objective}$), and, in contrast, requirements elicitations performed on rich media as a poor objective fit ($-_{objective}$).

Similarly, we consider the requirements negotiation task as a type of *equivocal* (Daft and Lengel 1986), intellective task that involves different sub-tasks, both *conflictual* and *behavioral* (McGrath 1984), and that requires a *high degree of member interdependence* (McGrath 1984), so that conflicts of interests are resolved to one shared solution. From a cognitive perspective, the communication process required to resolve ambiguities primarily involves *convergent thinking* (McGrath 1984) because seeking clarification and reaching closure reduces equivocality. Building then on the Social Presence theory (Short et al. 1976), one would assume that the convergent thinking process requires not only the exchange of task-oriented communication, but also of those *relational* cues, useful in resolving opposing individual views and in converging to a single shared view. In fact, task/technology fit theories consistently argue that using media that deliver less information than required by tasks generates poor fits due to effectiveness and quality problems (McGrath 1991). Media Synchronicity theory (Dennis et al. 2008) agrees that the reduction of *ambiguity* inherent in tasks for which the focus is on *convergence* can benefit from having rich communication channels available. Compared to lean media, rich media have a higher degree of *synchronicity*, which ensures *immediacy of feedback* during communication (Dennis et al. 2008), and convey a higher sense of *social presence* (Short et al. 1976), which in turn fosters *comfort* (Short et al. 1976) and individual *commitment* (i.e., motivation and attention) to task execution (Robert and Dennis 2005). In sum, the framework indicates requirements negotiations performed using rich media as a good objective task/technology fit ($+_{objective}$), whereas those performed on lean media as a poor objective fit ($-_{objective}$). Accordingly, from the second level of analysis we draw our second hypothesis:

H2 *TXT elicitation represents a better task/technology fit (*i.e.*, more effective) than F2F elicitation. F2F negotiation represents a better task/technology fit (*i.e.*, more effective) than TXT negotiation.*

In summary, the theories considered in our framework suggest that when objectively assessing task performance in TXT and F2F, TXT is a good fit in relation to the performance of requirements elicitation tasks, but a poor fit in relation to the performance of requirements negotiation tasks. However, because computer-mediated communication is the primary interest in this research, we also give special attention to the TXT communication medium itself. Knowledge about the affordances as well as challenges in using text-based communication for requirements tasks in the absence of F2F interaction is also important, especially in geographically distributed projects that may not benefit from mechanisms of interaction other than text-based communication.

Therefore, last but not least, we turn our attention to the direct comparison of the performance of elicitations and negotiations in TXT medium. Here, our framework indicates that, according to Common Ground Theory (Clark and Brennan 1991), achieving common ground in TXT negotiations appears to be more difficult than in TXT elicitations. According to Clark and Brennan (1991), in general, the grounding process changes with media and the cost of developing mutual understanding increases when using a lean text-based medium, due to the lack of attributes such as copresence, visibility, and audibility. Therefore, common ground is in general impaired in text-based communication, but the extent of such impairment, however, depends on tasks. Requirements negotiations are equivocal tasks during which achieving common ground relies on a high degree of member

interdependence to resolve ambiguities, reach agreement, and converge to shared ideas and solutions. In contrast, requirements elicitations do not require such a high degree of member interdependence to achieve common ground during activities such as generating ideas. Hence our third hypothesis:

H3 *The establishment of common ground is more impaired in TXT negotiations than in TXT elicitations.*

5 Empirical Study

In the following, we describe the details of our empirical study to test the hypotheses developed in the previous section. Based on the theoretical framework, our study investigates the support offered by a type of lean media, synchronous text-based communication (TXT), to requirements elicitations and negotiations, in comparison with face-to-face communication (F2F) as an example of rich media. In particular, we studied the performance of six groups of undergraduate students involved in a realistic software project whose outcome was a requirements specification (RS) developed through a process that involved specific requirements elicitation and negotiation tasks.

5.1 Study Design

We conducted the empirical study with six academic groups involved in an undergraduate RE course at the University of Victoria in 2006. The course involved thirty-eight students working in six project teams in the development of six realistic software projects (brief descriptions of the projects appear in Appendix 1). Selected by the students, the projects had to conform to a number of criteria meant to impose a certain level of similarity between the projects: implemented as web applications, they had a diverse range of stakeholders and end-users and entailed a significant amount of user-interaction. However, the task in our study was non-trivial. Unlike other experimental studies of requirements elicitations or negotiations, we designed our tasks for elicitations and negotiations of software requirements for non-toy-sized, realistic systems. Our study required the development and use of specialized knowledge from several different documents during the RE process.

Each team consisted of a client group and a developer group. The class was distributed into six groups (Gr1-6)—each group composed of five to eight randomly-selected students—and each group played the role of client group for one project team and the role of a developer for another project team simultaneously (see group assignment to projects in Table 1). For instance, students in group Gr1 acted as clients in Project1 and as developers in Project6.

Table 1 Groups and allocation to projects

Project	Client team	Developer team
Project1	Gr1	Gr2
Project2	Gr2	Gr3
Project3	Gr3	Gr4
Project4	Gr4	Gr5
Project5	Gr5	Gr6
Project6	Gr6	Gr1

The goal of each project team was to develop a Requirements Specification (RS) document as a negotiated software contract between the developer and client teams. Some coding by the developer groups was involved in the development of prototypes used in requirements validation activities. Figure 2 illustrates the workflow of the requirements development process followed by the teams over a period of about ten weeks. Each of the ten phases consists of tasks performed by either one of the client/developer groups or by both groups (joint project tasks). From the assigned project topic, each client group created a Request for Proposal (RFP) that invited the developers to propose their solutions to client needs. After developing the RFP document, the requirements elicitation activity was followed by the development of RS 1.0. Subsequently, a requirements negotiation activity followed an inspection of this document, results from which were incorporated in the RS 2.0 version.

The RS documents used the IEEE Std. 830–1998 template.[1] The deliverables on which students received a grade in the course were the RS 1.0 and 2.0, reflecting the common ground of the project that the clients and the developers built over the requirements elicitation and negotiation workshops.

To study our hypotheses about synchronous TXT and F2F communication, we focused our investigation on both the activities of requirements elicitation and negotiation in the course projects, and the outcome of the project, RS 2.0. We designed the study such that three project groups were involved in F2F elicitation followed by text-based negotiations, while the other three groups were involved in text-based elicitations followed by F2F negotiations (referred to as two process variants henceforth). Because each group of students was involved in two different projects playing two different roles, each student was in fact involved in the two process variants in different roles. Table 2 shows the experimental plan, which corresponds to a 2^3 factorial design (Mann and Whitney 1947). The three factors, each having two levels, are (1) *communication mode* (F2F and TXT); (2) *requirements task* (elicitation and negotiation); and (3) *role* (client and developer). The *communication mode* and *requirements* task factors varied within subjects, whereas the *role* factor varied between subjects.

5.2 Instrumentation, Training, and Execution

All the requirements workshop sessions were held in parallel and completed within one hour. While the F2F workshops were held as co-located meetings in the University laboratory, the students interacted from three different laboratories during the TXT sessions, so as to avoid F2F interaction. Each student was assigned to a given seat to prevent entire teams co-located in the same laboratory, or some participants in the same workshop sitting side by side. According to the course design, each of the F2F and TXT requirements elicitation sessions involved two developers and the entire client team, whereas F2F and TXT negotiations involved the entire project teams (i.e., all clients and developers together). The text-based requirements elicitations and negotiations were run using the eConference tool (see Fig. 3), a synchronous text-based, distributed meeting system (Calefato and Lanubile 2009). The primary functionality provided by the tool is a closed group chat, augmented with collaborative features that help to control distributed meetings. The message board allows each participant to see the messages sent in via the input panel.

The agenda indicates the status of the meeting, as well as the current item under discussion. The edit panel synthesizes a summary of the discussion. The presence panel shows the participants currently logged in and the role they play. Finally, the hand-raise

[1] http://standards.ieee.org/reading/ieee/std_public/description/se/830-1998_desc.html

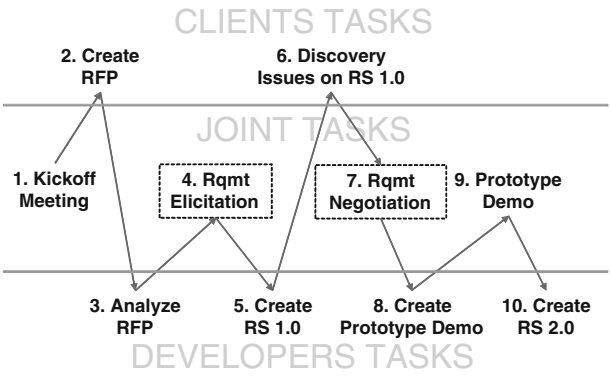


Fig. 2 Workflow for the development process of the requirements specification documents

panel mimics the hand-raise social protocol that people use during real meetings to coordinate discussion and turn-taking. To allow participants to gain familiarity with the tool, they received a 1-hour demonstration at class time, a training session 1 week before each TXT workshop session, and a user manual on the course web site.

During the TXT workshops, one researcher, a teaching assistant, and a Ph.D. student were available in the laboratory to provide technical support and ensure lack of F2F interaction outside the system. It was important to the study that the participants in the text-based sessions did not have access to any visual or verbal cues, typically unavailable in text-based communication.

5.3 Data Collection

We collected data using three sources and instruments: 1) post-elicitation and post-negotiation questionnaires; 2) six interaction logs collected from the TXT elicitations and the TXT negotiations; and 3) the RS 2.0 documents.

The two post-elicitation and post-negotiation questionnaires were administered to the students in both electronic and printed form about 1 week after each requirements workshop session. We had 17 out of 24 students respond to the post-elicitation questionnaire (response rate 71%), and 19 out of 38 students respond to the post-negotiation questionnaire (response rate 50%). In both cases, respondents were evenly distributed among the groups.

Table 2 The 2^3 factorial design of the experiment

	A Comm. Mode	*B* Rqmt Workshop	*C* Role	Subjects
(1)	F2F	elicitation	client	Gr1,Gr3,Gr5
a	TXT	elicitation	client	Gr2,Gr4,Gr6
b	F2F	negotiation	client	Gr2,Gr4,Gr6
ab	TXT	negotiation	client	Gr1,Gr3,Gr5
c	F2F	elicitation	dev	Gr2,Gr4,Gr6
ac	TXT	elicitation	dev	Gr1,Gr3,Gr5
bc	F2F	negotiation	dev	Gr1,Gr3,Gr5
abc	TXT	negotiation	dev	Gr2,Gr4,Gr6

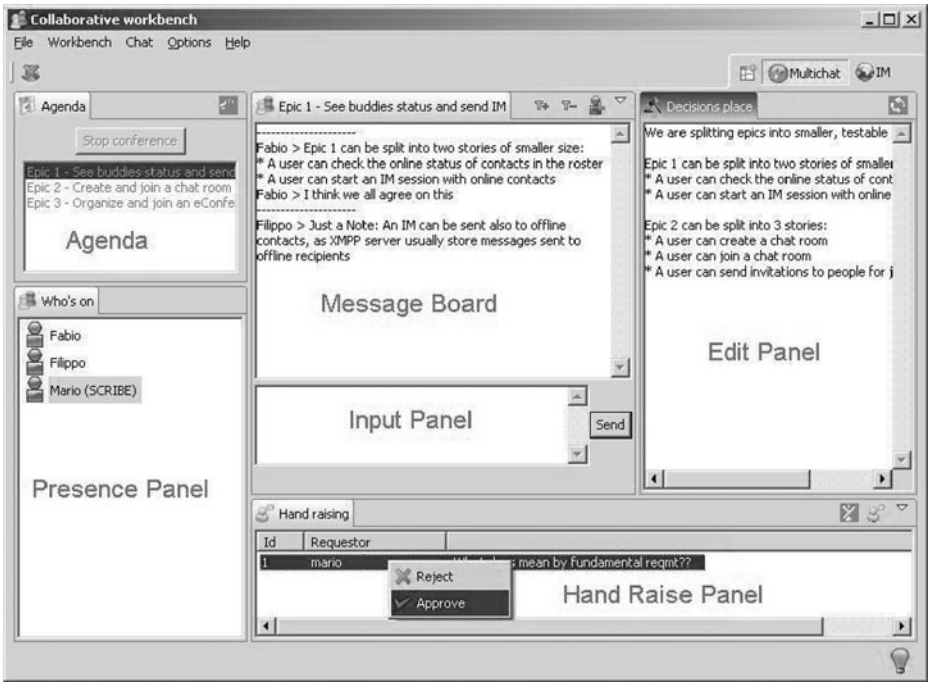


Fig. 3 A screenshot of eConference

The questionnaires were formulated taking into account the communication problems in RE (Al-Rawas and Easterbrook 1996) and the issues informally reported by the students after each requirements workshop session. The questionnaires contained both multiple choice and Likert scales without a central category to avoid the central tendency bias and to mitigate the social desirability bias (Cummins and Gullone 2000; Garland 1991).

At the end of the six requirements workshops conducted remotely through the eConference meeting tool (i.e., three TXT elicitations and three TXT negotiations), a log was automatically stored, containing all timestamped messages exchanged by the stakeholders during each meeting.

5.4 Dependent Variables, Measurements, and Results

Here, we describe the constructs and measurements that we defined to conceptualize the elements in our hypotheses.

5.4.1 Comfort with Communication Mode and Satisfaction with Performance in F2F and TXT Requirements Workshops

We remind the reader of our first study hypothesis:

H1 *F2F requirements elicitations (F2F negotiations) are better appreciated than TXT requirements elicitations (TXT negotiations).*

In our study, we conceptualize the level of *appreciation* as the level of (1) *comfort with communication mode* and (2) *satisfaction with performance*.

Thus, **H1** is rephrased as follows:

> **H1a′**—*The perceived level of comfort with the communication mode in F2F elicitation (F2F negotiation) is higher than in TXT elicitation (TXT negotiation).*
> **H1b′**—*The perceived level of satisfaction with performance in F2F elicitation (F2F negotiation) is higher than in TXT elicitation (TXT negotiation).*

To measure the *comfort with communication mode* construct, we used a 5-item 4-point Likert scale (4=strongly agree and 1=strongly disagree) to assess the perceived degree of discussion contentment and engagement level in each of the four workshops (see Table 3).

We also employed a scale that included a number of multiple choice items to assess aspects such as motivation, domination, and ability to express and understand ideas (see Tables 4 and 5).

To measure *satisfaction with performance*, we used a similar 4-point Likert scale to assess participant perception of whether decisions were consensus based and of whether the amount of information generated was properly processed in both F2F and TXT workshops. Namely, the two questions were Q27 *"During the workshop, decisions were made with a group consensus*" and Q28 *"During the workshop, the discussion was too fast and information was missed."* We also utilized a scale that included a number of multiple-choice items to assess the support for structured or in-depth discussions (see Tables 6 and 7).

We used nonparametric statistics in our analysis because the sample was rather small and we could not rely on the normality assumption. First, to ensure the validity of the construct of *comfort with communication mode*, we applied principal components analysis (Hatcher and Stepanski 1994) to the five related questions shown in Table 3, which appeared in both the post-elicitation and post-negotiation questionnaire. The procedure extracted the same component from the scale in both the elicitation and the negotiation questionnaire. Three items were retained in the extracted component (shown in Table 8 with the breakdown of participant responses) and two were discarded.

To test **H1a′** and identify differences in the perceived level of *comfort with communication mode*, we performed chi-square goodness of fit tests on the data from the three extracted components (see Table 8) and the data from the multiple choice questions listed in Tables 4 and 5. On the one hand, F2F was preferred for requirements elicitations due to its support to *facilitate familiarization with other participants* (see Table 4, Q12, χ^2=13.24, p=.000), *better support the ability to express complex ideas* (Q10, χ^2=23.06, p=.000), and *understand others' opinions* (Q13, χ^2=7.12, p=.000). Similarly, F2F was preferred for requirements negotiations to *facilitate familiarization with other participants* (see the Table 5, Q21, χ^2= 18.11, p=.000), *better support the ability to express complex ideas* (Q19, χ^2=9.58, p=.008),

Table 3 Questions asked to evaluate comfort with communication mode during elicitation and negotiation requirements workshops

In comparison to F2F workshops…
1. TXT workshop offered increased opportunities to participate in the discussion
2. TXT workshop encouraged participants to more openly discuss conflicting issues with members of their own group
3. TXT workshop meeting encouraged participants to more openly discuss conflicting issues with members of the other group
4. TXT workshop needs/requires a higher level of preparation
5. TXT workshop grants stakeholders a higher level of comfort (lower pressure felt, making comments without being afraid of intimidation)

Table 4 Questions asked about comfort of interaction during F2F and TXT requirements elicitations and results of the goodness of fit test (N=17)

Which medium better supported…		F2F	TXT	About the same	Chi-square χ^2
Q7	*articulation of ideas freely*	7	5	5	.47
Q8	*spontaneous discussion*	10	3	4	5.06
Q9	*ability to express basic ideas*	7	4	6	.82
Q10	*ability to express complex ideas*	**15**	1	1	**23.06**
Q11	*control of etiquette and professionalism*	7	5	5	.47
Q12	*gaining familiarity with participants*	**16**	–	1	**13.24**
Q13	*ability to understand participants' opinions*	**14**	0	3	**7.12**
Q14	*sense of involvement in the workshop*	6	2	9	4.35
Q15	*motivation to participate in discussions*	4	3	10	5.06
Total		86	23	44	

Significant results at the 5% level are shown in bold

understand others' opinions (Q22, χ^2=11.79, p=.002), and *control of etiquette and professionalism* (Q20, χ^2=11.84, p=.001).

On the other hand, the test results in Table III show that TXT is also preferred for other reasons than those listed above. Compared to F2F requirements workshops, TXT elicitations encourage participants to *more openly discuss conflicting issues with same and other group members* (χ^2=11.48, p=.009, and χ^2=9.12, p=.028). Similarly, TXT was preferred for negotiations because of an increased *opportunity to participate in the discussion* and ability to *more openly discuss conflicting issues with same group members* (χ^2=10.68, p=.014, and χ^2=8, p=.018).

To note is that these statistics compare responses from the four workshops without special regard to the role played by the participants in the workshops. Hence, we performed the Mann–Whitney U test (Conover 1980), a nonparametric alternative to t-test on independent samples, to verify whether the role factor (client or developer) influenced

Table 5 Questions asked about comfort of interaction during F2F and TXT requirements negotiations and results of the goodness of fit test (N=17)

Which medium better supported…		F2F	TXT	About the same	Chi-square χ^2
Q16	*articulation of ideas freely*	6	5	8	.74
Q17	*spontaneous discussion*	6	5	8	.74
Q18	*ability to express basic ideas*	7	2	10	5.16
Q19	*ability to express complex ideas*	**12**	1	6	**9.58**
Q20	*control of etiquette and professionalism*	**17**	–	2	**11.84**
Q21	*gaining familiarity with participants*	**15**	1	3	**18.11**
Q22	*ability to understand participants' opinions*	**13**	1	5	**11.79**
Q23	*sense of involvement in the workshop*	8	5	6	.74
Q24	*motivation to participate in discussions*	6	5	8	.74
Q25	*look up relevant info from existing documents*	6	8	5	.74
Q26	*avoiding somebody to dominate discussion*	2	10	7	5.16
Total		98	43	68	

Significant results at the 5% level are shown in bold

Table 6 Questions asked about satisfaction with performance during F2F and TXT requirements elicitations and results of the goodness of fit test (N=17)

Which medium better supported…		F2F	TXT	About the same	Chi-square χ^2
Q1	*consensus-based decisions*	7	6	4	.82
Q2	*structured discussion*	1	**13**	3	**14.59**
Q3	*documentation of decisions made*	1	**13**	3	**14.59**
Q4	*workshop facilitation*	0	12	5	2.88
Q5	*in-depth discussion*	**12**	2	3	**10.71**
Q6	*visibility of decisions made*	–	12	5	2.88
Total		21	58	23	

Significant results at the 5% level are shown in bold

participant perception of comfort with the communication mode in both paired comparisons. We found no significant difference, indicating that the respondent role did not affect the perception of comfort.

To determine significant differences in the level of participant *satisfaction with performance*, first we analyzed the responses to questions Q27 (*"During the workshop, decisions were made with a group consensus*") and Q28 (*"During the workshop, the discussion was too fast and information was missed"*). We calculated per-subject satisfaction scores by summing all responses for each respondent to obtain an overall score of the personal level of satisfaction with performance during the requirements workshops. Then, the ranks of the workshop/medium fits were calculated based on the per-subject summed scores. The two box plots in Fig. 4 show that the difference in mean scores of the personal level of satisfaction with performance between F2F elicitation (2.75) and TXT elicitation (2.15, see Fig. 4a) is not as large as in the case of F2F negotiation and TXT negotiation, which, respectively, exhibit the highest (3.5) and the lowest (1.6) average ranks (see Fig. 4b).

Table 7 Questions asked about satisfaction with performance during F2F and TXT requirements negotiations and results of the goodness of fit test (N=19)

Which medium better supported…		F2F	TXT	About the same	Chi-square χ^2
Q1	*consensus-based decisions*	5	7	7	.42
Q2	*structured discussion*	3	11	5	5.47
Q3	*documentation of decisions made*	1	**13**	5	**11.79**
Q4	*workshop facilitation*	–	**16**	3	**8.9**
Q5	*in-depth discussion*	**12**	2	5	**8.32**
Q6	*visibility of decisions made*	–	**17**	2	**11.84**
Q7	*keeping participants on task*	10	6	3	3.9
Q8	*reaching an agreement*	7	4	8	1.37
Q9	*prioritizing requirements*	5	5	9	1.68
Q10	*resolving conflicts*	10	–	9	.05
Q11	*drawing conclusions*	10	5	4	3.26
Q12	*getting all the work done*	8	2	9	4.53
Total		71	88	69	

Significant results at 5% level are shown in bold

Table 8 Evaluation of comfort with communication mode and results from the chi-square goodness of fit test (only the items extracted by principal component analysis are listed)

"In comparison to F2F workshops, TXT workshops…"	1. "offered increased opportunity to participate in the discussion"				2. "encouraged to more openly discuss conflicting issues with same group members"				3. "encouraged to more openly discuss conflicting issues with other group members"			
	Elicitation		Negotiation		Elicitation		Negotiation		Elicitation		Negotiation	
	N	%	N	%	N	%	N	%	N	%	N	%
Strongly agree	5	29.4	1	5.3	2	11.8	2	10.5	5	29.4	–	–
Somewhat agree	6	35.3	**10**	**52.6**	**10**	**58.8**	8	42.1	**9**	**52.9**	**11**	**57.9**
Somewhat disagree	4	23.5	6	31.6	4	23.5	6	31.5	2	11.8	7	36.8
Strongly disagree	3	11.8	2	10.5	1	5.9	3	15.8	1	5.9	1	5.3
Chi-square χ^2	2.06		**10.68**		**11.48**		4.79		**9.2**		**8**	

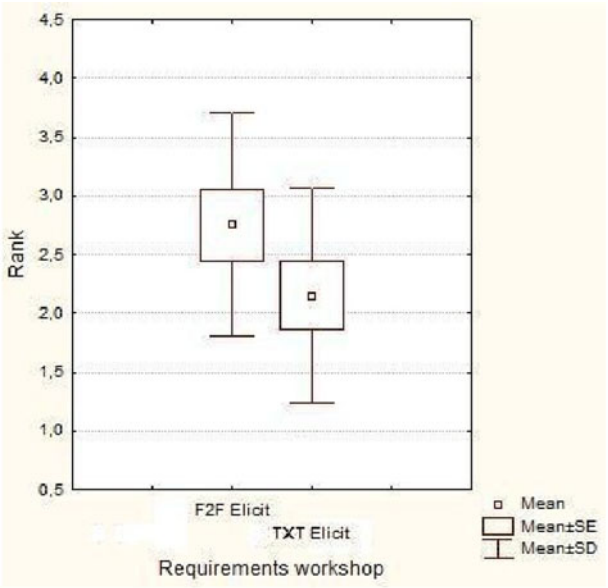
Significant results at the 5% level are shown in bold

To test **H1b′** and identify differences in the perceived level of *satisfaction with performance*, we applied a series of statistics to these scores to perform matched-pair comparisons between (I) F2F elicitation vs. TXT elicitation, and (II) F2F negotiation vs. TXT negotiation. We performed the comparison by applying the Wilcoxon signed-rank test, as a nonparametric alternative to the *t*-test for two dependent samples (Conover 1980). The results, displayed in Table 9, show for each matched-pair comparison, positive ranks (i.e., how many participants preferred F2F elicitation or negotiation to the TXT counterpart), negative ranks (i.e., how many participants preferred TXT elicitation to the F2F counterpart), and ties (i.e., how many participants perceived F2F and TXT workshops to be equal). The two Wilcoxon tests indicate that, while participants significantly preferred F2F negotiation to TXT negotiation ($Z=2.54$, $p=.011$), no statistically significant difference was found in the comparison between F2F elicitation and TXT elicitation ($Z=1.56$, $p=.119$).

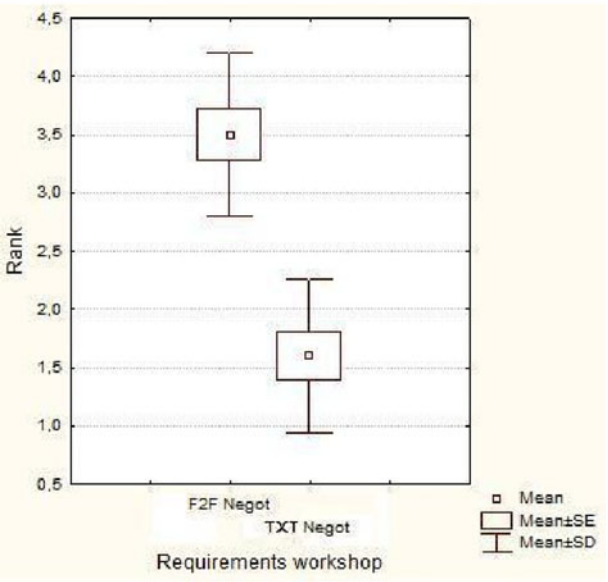
The second analysis of the data on *satisfaction with performance* was a number of chi-square goodness of fit tests on the responses to the questions shown in Tables 6 and 7. With respect to requirements elicitations, the F2F communication mode was preferred for its support for *in-depth discussion* (see Table 6, Q5, $\chi^2=10.71$, $p=.001$). Conversely, a text-based communication mode was preferred for its better support for *structured discussions* (Q2, $\chi^2=14.59$, $p=.001$) and for *documentation of decisions made* (Q3, $\chi^2=14.59$, $p=.001$). Finally, with respect to requirements negotiations, F2F was again preferred for having an *in-depth discussion* (see Table 7, Q5, $\chi^2=8.32$, $p=.003$). The text-based communication mode was instead preferred not only for better *documentation of decisions made* (Q3, $\chi^2=11.79$, $p=.003$), but also for better *workshop facilitation* (Q4, $\chi^2=8.9$, $p=.016$) and *visibility of decisions made* (Q6, $\chi^2=11.84$, $p=.001$).

5.4.2 Objective Assessment of Effectiveness of F2F and TXT Requirements Workshops

With our second hypothesis, we sought to obtain an objective assessment of the effectiveness and group performance in the TXT and F2F requirements sessions, and



a) F2F vs. TXT elicitations



b) F2F vs. TXT negotiations

Fig. 4 Ranks based on subjects' evaluation of satisfaction with performance in F2F vs. TXT elicitation (**a**) and F2F vs. TXT negotiations (**b**)

assumed that the higher the effectiveness the better the task/technology fit. The hypothesis, as stated above, is:

H2 *TXT elicitation represents a better task/technology fit (*i.e.*, more effective) than F2F elicitation. F2F negotiation represents a better task/technology fit (*i.e.*, more effective) than TXT negotiation.*

Table 9 Results from the Wilcoxon signed-rank test for the matched-pair comparisons

Matched-pair comparison (A vs. B)	Positive ranks (A>B)	Negative ranks (A<B)	Ties (A=B)	Wilcoxon test
I. F2F elicit vs. TXT elicitat	5	2	3	Z=1.56
II. F2F negot vs. TXT negot	8	0	2	**Z=2.54**

Results significant at the 5% level are shown in bold

In our study, we conceptualized the objective assessment of effectiveness as the number of defects inserted in the final RS document produced by each group. As shown in Fig. 5, the two factors considered in this study, that is, *requirements workshop* and *communication mode*, created two variants in the requirements definition process. Three of the six projects used TXT Elicitation followed by F2F Negotiation (referred to as the TE-FN process variant henceforth). The other three projects used F2F Elicitation followed by TXT Negotiation (the FE-TN process variant). Hypothesizing that TXT elicitation and F2F negotiation are, respectively, better task/technology fits than F2F elicitation and TXT negotiation implies that the TE-FN process variant is more effective than the FE-TN process variant. Accordingly, **H2** is revised as follows:

> **H2′** – *The performance of groups following the TE-FN process variant is higher than that of groups following the FE-TN process variant.*

An independent and distributed team of three inspectors inspected the RS 2.0 documents produced as the outcome of each of the six projects. Thus, the *group performance* dependent variable was operationalized in terms of the number of defects found by the inspection team, the number of major defects, and the defect density (i.e., the number of defects per page). In particular, the inspection team members first inspected each document independently, then they merged and discussed the documents until they obtained a shared list of defects per project.

To test our **H2′** hypothesis about group performance in the two process variants, we conducted a formal inspection of the six RS 2.0 documents. As a result, of the inspection, we compared the number of defects found by the inspection team for each document produced in the two process variants (see Table 10 for summary statistics per process variant). Again, because of the small number of observations (6 projects/documents) and the impossibility of relying on the normality assumption, we performed the Mann–Whitney *U* test (Conover 1980). The test failed to reveal any significant effect for the process variant (U=2, p=.2). The results do not change if we consider as the metric of performance the number of major defects found or the defect density. In fact, for the test to be significant, all the three projects executing one process variant should outperform all the remaining three projects executing the other variant (Mann and Whitney 1947). In this case, however, Project3 following the FE-TN variant outperformed (i.e., only 19 defects) two of the three projects that executed the TE-FN variant, namely Project2 and Project6 (21 and 31 defects, respectively). Hence, no evidence supports hypothesis **H2′**. Finally, we point out that one of the researchers reformatted all the RS documents to match the same template in computing correctly the defect density metric.

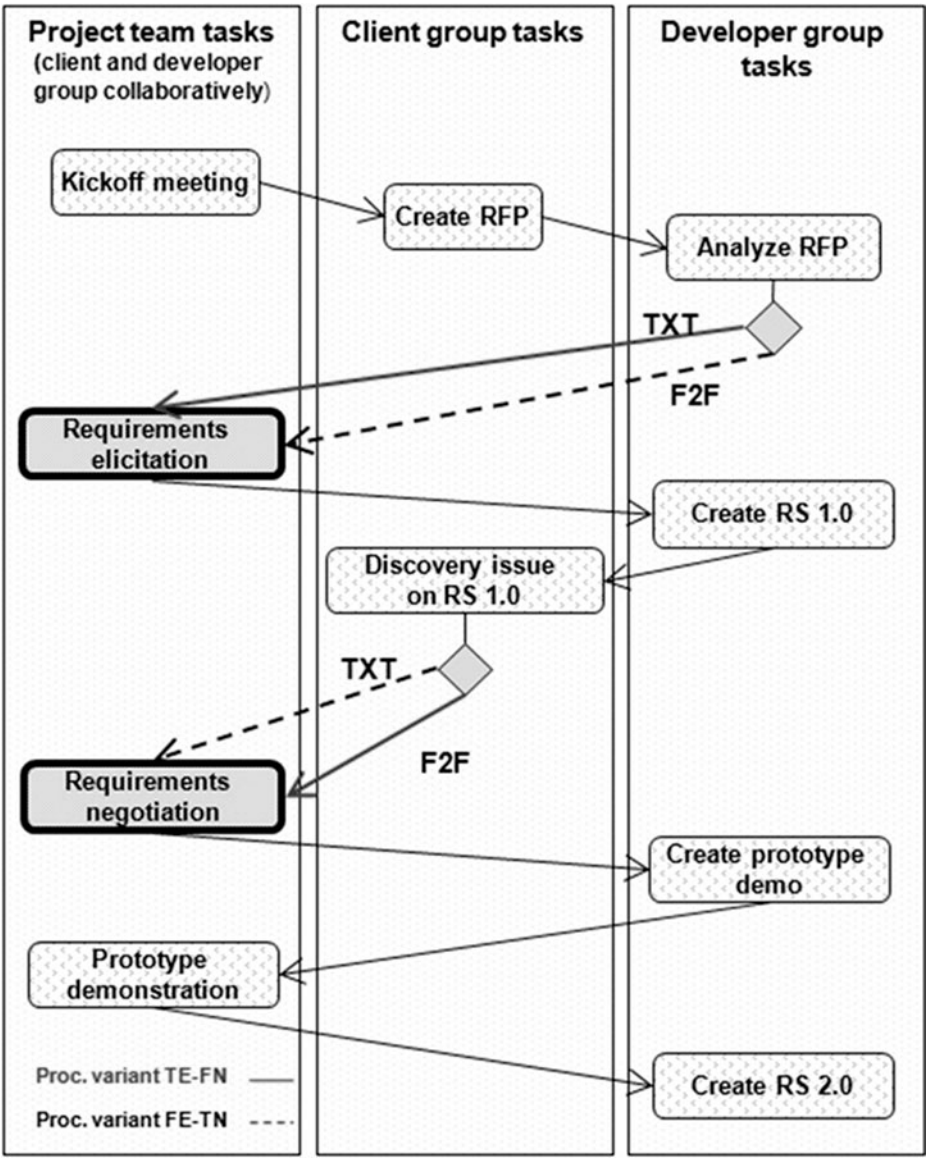


Fig. 5 The two variants in the requirements definition process

5.4.3 Common Ground Impairment in TXT Requirements Workshops

To gain more insight on the effectiveness of TXT requirements workshops, we looked at the basic goal of communication, that is, establishing a common ground. We remind the reader of our third study hypothesis:

H3 *The establishment of common ground is more impaired in TXT negotiations than in TXT elicitations.*

Table 10 The number of defects, major defects, and defect density found from the inspection of the final RS 2.0 documents

Project (Clients/Dev)	Process Variant	Defects	Major Defects	# Pages RS 2.0	Defect Density[a]
Project 1 (GR1/GR2)	FE-TN (Negot A)	43	24	39	1.10
Project 2 (GR2/GR3)	TE-FN (Elicit B)	21	9	32	0.66
Project 3(GR3/GR4)	FE-TN (Negot C)	19	8	28	0.68
Project 4 (GR4/GR5)	TE-FN (Elicit D)	17	7	25	0.68
Project 5 (GR5/GR6)	FE-TN (Negot E)	36	22	25	1.44
Project 6 (GR6/GR1)	TE-FN (Elicit F)	31	17	34	0.91

[a] computed as the number of defects found per page

Measuring the level of common ground that groups achieve is generally a challenging task (Clark and Brennan 1991). However, to confirm our hypothesis about common ground impairment, we did not need to measure directly and to compare the levels of common ground achieved during TXT elicitations and negotiations. Instead, we may collected evidence of *lack of common ground* during TXT elicitation and negotiation workshops (Clark and Brennan 1991). Thus, the third hypothesis **H3′** is given as follows:

> **H3′** – *The lack of common ground manifested by stakeholders in TXT elicitations is lower than in TXT negotiations.*

We operationalized the construct of lack of common ground in terms of *negative evidence* and *grounding chain*, defined as follows. Receivers provide negative evidence during communication when messages are improperly or incompletely understood. In other words, the more complex the task (e.g., ambiguous, equivocal), the higher the number of ill-defined messages presented by speakers and, consequently, the more negative evidence presented by receivers. Questions can also be a means for grounding (i.e., asking for further clarification or information helps establish common ground). In addition, according to Daft and Lengel (1986), in equivocal situations "participants are not certain about what questions to ask and, if questions are posed, the situation is ill-defined to the point where a clear answer will not be forthcoming" (p. 556–557). In other words, questions tend to snowball in the sense that an initial question causes several subsequent questions for clarification and their answers, as well as the presentation of other utterances (e.g., requests) eliciting further information before a common ground is reached. We call this sequence a *grounding chain* because it builds on the concept of "chaining rule" identified in previous conversation analysis literature (McHoul and Rapley 2001; Titscher et al. 2000). The question-answer chain is a conversational mechanism referred to as an "adjacency pair" (e.g., invitation-acceptance, greeting-return), typical of two-party conversation. Two-party conversations, including interrogations or interviews, allow the pursuit of information from others, its assembly, and conclusions through a series of questions (Ten Have 1999). Although formulated with respect to F2F conversation, adjacency pairs can be identified also in text-based chat, but with one exception due to parallel input, which makes turn-taking organization more difficult. In text-based chat, adjacent pairs are often broken. Because typing is slower than speaking, presenters tend to break long utterances into several pieces so receivers do not have to wait too long to read the whole message, but allowing pieces from different utterances to get interleaved with each other, thus breaking adjacency pairs.

To test **H3′** about common ground in TXT sessions, we investigated the presence of negative evidence and the length of grounding chains. To quantify presence of negative evidence and grounding chains we performed content analysis on the logs of the TXT workshops.

Content analysis, also called coding (Stemler 2001), is a mix of quantitative and qualitative analysis (Seaman 1999; Taylor and Bogdan 1984), which transforms qualitative data (i.e., free text, spoken or even written, as in our case) into quantitative data. Unlike in qualitative research, in content analysis data are classified according to a finite set of possible categories, so the analysis may be readily quantifiable. To perform such analysis, we defined a coding schema that satisfied our specific need to look at the lack of common ground.

The nine categories, or thematic units, we identified from the logs of the six TXT requirements workshops are listed in Appendix 2. The units that relate to common ground are *questions*, *answers* (both further categorized as *yes-no* and *complex*), *agreements*, *disagreements*, *checks* (further categorized as *provisional*, *verbatim copy*, and *misunderstanding*), and *acknowledgements*.

Two coders, one of the researchers and a research assistant, performed the content analysis separately and the intercoder agreement, measured by Cohen's Kappa, ranged from 0.84 for Negotiation E (Project 5 - GR5/GR6), to 0.93 for Negotiation C (Project 3 - GR3/GR4). Table 11 shows the details of the conversation logs for each of the six TXT requirements workshops, that is, the breakdown of the thematic units identified for each category related to the measure of common ground (for the sake of space we report only those categories that directly contribute to the measure of lack of common ground).

To identify grounding chains, we searched for the units that would initiate a grounding chain, that is, *disagreements*, *questions,* and *misunderstanding checks*, and provide evidence of lack of common ground (i.e., information available is equivocal or incomplete). A grounding chain is closed by an utterance categorized as *answer*, *agreement*, or *acknowledgment* (i.e., information available is clear and complete). We define the length of a grounding chain as the end-to-end thematic unit count from an opening to a conclusion of a grounding chain. The table in Appendix 3 provides a short example of one grounding chain identified in one of the TXT workshop logs.

To quantify negative evidence, we counted all the *disagreements* and those *checks* that expressed misunderstanding in each workshop. The negative evidence per grounding chain was computed as the overall amount of negative evidence from a workshop divided by the number of grounding chains in the same workshop. Such normalization is necessary when comparing meetings due to the large differences between the lengths (i.e., overall number of units) of the six workshops (range 145–663 units).

Once operationalized, the **H3′** hypothesis becomes:

> **H3a′** – *Grounding chains are shorter in elicitations than in negotiations.*
> **H3b′** – *Negative evidence per grounding chain is lower in elicitations than in negotiations.*

To evaluate **H3a′** we first computed the lengths (expressed as the number of thematic units) of the grounding chains for each of the six requirements workshops. We wanted to determine differences in the lengths of the grounding chains between both the six workshops and the two types of workshops (i.e., elicitations and negotiations). This design corresponds to an unbalanced two-stage nested design where the *type of workshop* is the fixed factor and the *workshop* is the nested (or random) factor (Montgomery 1996).

Table 11 The results of applying the coding schema to the TXT workshop logs

	# of Thematic Units	# of Grounding Chains	# QUESTIONS	# DISAGREEMENTS	# CHECKS Misunderstandings (Total)	Negative Evidence	Amount of Negative Evidence per Grounding Chain
Elicit B	663	24	62	21	18 (59)	39	1.63
Elicit D	145	9	17	7	4 (11)	11	1.22
Elicit F	172	7	21	7	5 (11)	12	1.71
Negot A	347	12	43	23	10 (22)	33	2.75
Negot C	500	5	35	31	15 (38)	46	9.2
Negot E	194	6	22	8	9 (17)	17	2.83

The data on the lengths of the grounding chains identified in the six workshops are shown as a box plot in Fig. 6. However, the data does not lend itself to analysis of variance because the assumption about residuals is not respected (Montgomery 1996). More precisely, the variability within the negotiation workshops is consistently larger than in the elicitation workshops and results from tests such as analysis of variance are not conclusive.

The last column of Table 11 shows the amount of negative evidence per grounding chain for each of the elicitation and negotiation workshops. To evaluate the **H3b′** hypotheses we performed the Mann–Whitney *U* test (Montgomery 1996) and found that the amount of negative evidence per grounding chains in the elicitation workshops can be considered significantly lower than in the negotiation workshops ($Z=1.964$) at the 0.1 level.

6 Discussion

The main contribution of this paper is the empirical evidence from a study that tested three hypotheses developed from our framework of analysis of group performance in synchronous text-based distributed requirements elicitations and negotiations. The theories of media richness and task/technology fit reviewed here indicate that no trivial answer is available regarding which medium is best suited for effective performance of communication-intensive tasks in RE.

Contrary to common belief and predictions drawn from the theories in the framework, in our study the rich F2F medium was not always the *preferred* medium for requirements tasks. Our study findings indicate aspects of requirements workshops for which the text-based communication was more beneficial. These aspects include support for structured discussion, proper documentation, and visibility of decisions made, as well as increased ability to discuss openly conflicting issues. Moreover, our objective evaluation of group performance specific to computer-mediated requirements elicitations *and* negotiations suggests that the development of common ground, essential to achieving shared understanding of requirements, is significantly more difficult in computer-mediated negotiations than in elicitations, speaking further about the benefits of computer-mediated

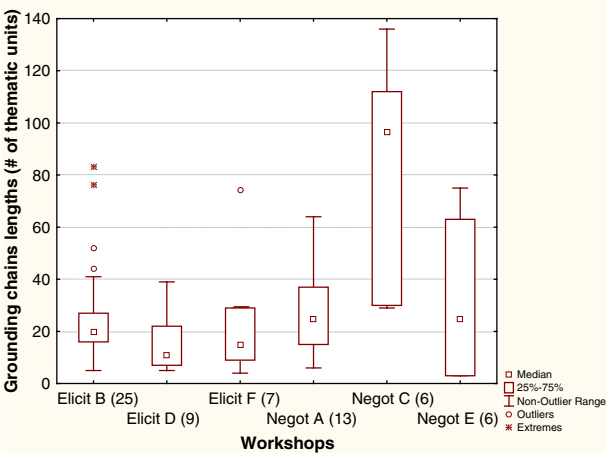


Fig. 6 The lengths of the grounding chains identified in each of the six TXT requirements workshops (numbers in brackets show the count of grounding chains)

requirements elicitations. Our findings add to the evidence about computer-mediated communication in RE, other than what we know about audioconferencing (Lloyd et al. 2002) and videoconferencing (Damian et al. 2003) support for requirements tasks, and provide some guidance to practitioners regarding the communication medium to choose to avoid losses in requirements task performance.

6.1 No Conclusive Evidence that F2F is the Preferred Medium for Requirements Tasks

Based on this framework, in our empirical study we tested the proposition that F2F communication was a good subjective fit and TXT a poor fit for both requirements elicitations and negotiations by comparing participant appreciation of the F2F- and TXT-based requirements workshops. We conceptualized the level of appreciation of the requirements workshops in terms of the perceived levels of *comfort with communication mode* and *satisfaction with performance*. Given the high synchronicity in the F2F medium, as well as its ability to foster commitment and to support relational communication, we expected that the respondents would prefer F2F communication to text-based communication for *both* the elicitation and negotiation tasks, thus rating F2F higher for both comfort and satisfaction with performance.

However, our findings show mixed evidence with respect to the *comfort with communication mode.* In both requirements elicitations and negotiations, we found that the F2F medium provided respondents with more opportunity to familiarize with others, and provided an enhanced ability to express complex ideas and to understand others' opinions. Interestingly, text-based communication was rated higher than F2F communication on the ability to discuss openly conflicting issues. These results confirm the predictions of socio-psychological theories that the depersonalization effect induced by the use of less-rich and less-social media limits domination, group/social pressure, and other dysfunctional aspects intrinsic to F2F group communication (Walther 1996) and that are specific to requirements group approaches (Gottesdiener 2002; Macaulay 1996).

With respect to *satisfaction with performance*, our findings suggest a different preference for F2F communication depending on the task. Contrary to our expectation that a F2F medium will rate higher on satisfaction than text-based communication when supporting *both* elicitations and negotiations, we found only that respondents were more satisfied with the performance in the F2F than in the TXT negotiations. We found no difference in the perceived satisfaction with performance between F2F and TXT elicitation*s*. These results suggest that the general preference for F2F requirements workshops may be due to the strong preference for the F2F negotiation fit over the TXT counterpart and not for the F2F medium in general. Further, regardless of the type of requirements workshop, we found that either communication medium is preferred depending on specific aspects of satisfaction with performance. The F2F medium was preferred for its support to build deep discussions, whereas TXT was preferred for having better task facilitation specifically in structured discussion, proper documentation, and visibility of decisions made. These results complement findings from previous GSS-related research that groups interacting on text-based channels often outperform collocated groups in tasks of idea generation because of the ability to input ideas in parallel (Fjermestad and Hiltz 1998; Murthy and Kerr 2000). In RE in particular, these findings suggest that groups having to elicit requirements in lean media do not necessarily underperform in comparison with groups in the traditional face-to-face elicitation workshops, and that text-based communication can be a better choice of medium for such tasks.

6.2 Communication Medium Might not Affect Group Performance

Our study also tested two hypotheses about the objective communication medium fit to requirements elicitations and negotiations. The first objective measure of performance was the number of defects in the requirements specification documents, as they reflected the shared understanding and decisions made during the requirements elicitations followed by negotiation sessions.

We hypothesized that groups using TXT elicitation followed by F2F negotiation (TE-FN process variant) would outperform the groups using F2F elicitation followed by TXT negotiation (FE-TN process variant). In other words, we tested whether using a combination of a good task/technology fit in the requirements definition process (TXT elicitation and F2F negotiation) would result in fewer defects in the RS document.

We did not find support for this hypothesis. As shown by the number of defects found in the six RS documents, not all three projects that used the TE-FN variant outperformed the three other projects following the FE-TN variant (and which would have resulted in statistical significant results). This result is also surprising since one would expect that using either F2F negotiation or TXT negotiation (indicated as good and poor task/technology fit, respectively, by the framework of analysis and perceived similarly by the participants in the study) would steer the group performance in a particular direction. Our study, however, included a small sample of groups and future research should study this conceptualization of objective performance with a larger sample where statistical significance is not as heavily reliant on groups in one experimental condition outperforming groups in the other condition.

6.3 Text-based Elicitations Offer Support to Achieving Common Ground

The second objective measure of performance in our study was the achievement of common ground, which allowed us to compare the objective fit of synchronous text-based communication during requirements elicitations with that of requirements negotiations. Specifically we hypothesized that groups would achieve more common ground (shorter grounding chains and less negative evidence per grounding chain) in TXT elicitations than in TXT negotiations. The higher the ambiguity in the task (e.g., requirements negotiations), the less common ground could be achieved in computer-mediated communication. Hence, we expected that the grounding chains would be shorter in TXT elicitations than in negotiations; as well, we expected the amount of negative evidence per grounding chains to be lower in the TXT elicitations than in negotiations. The data we analyzed allowed us to support the second part of this hypothesis, specifically that the TXT elicitations had less negative evidence than the TXT negotiations, confirming our expectations that common ground was more difficult to achieve in the TXT negotiations. Negative evidence, as an indication of lack of common ground in conversations, was identified when participants expressed disagreements or misunderstanding during the workshop. Thus, our findings suggest that the TXT groups converged to common ground faster in the elicitation than in the negotiation workshops. Unfortunately the high variability in our grounding chains data made the results of any analysis of variance unreliable (something we could not have anticipated *a priori*) and us unable to conclude the first part of our hypothesis on common ground.

These findings—that lean media offer less support to achieving common ground during requirements negotiations than during elicitations—add to the previous findings that TXT negotiations represent a poor task/technology fit. From a practical standpoint, they suggest

that when distributed groups rely on computer-mediated communication in performing requirements tasks, requirements elicitations, rather than negotiations, represent a less risky choice when it comes to achieving shared understanding of requirements. The design of computer-mediated tools to support RE tasks should focus on the particular needs of the elicitation tasks in distributed teams since they offer most opportunity for successful application.

6.4 Threats to Validity

One of the key issues in experimentation is evaluating the validity of results (Wohlin et al. 2000). Here, we discuss threats to validity of our findings.

6.4.1 Threats to Conclusion Validity

In our study, conclusion validity concerns arise due to our small sample size. To mitigate this threat, we used nonparametric tests in the statistical analysis, which do not rely on any assumed distribution of the underlying data and can be valid for even a small sample size. Empirical studies in software engineering are often performed with a low number of subjects. Small sample size often results in studies with low power, considered a potentially strong contributor for not finding statistical significance of empirical results (Conover 1980). Ideally, we should have performed a power analysis before conducting our study to ensure that the experimental design could find a statistically significant effect if one exists. However, in our case, an *a priori* power analysis was problematic because the effect size was unknown.

We adopted a 10% significance level to draw conclusions when testing for differences between the negative evidence in TXT elicitations and TXT negotiations. This is because a more relaxed alpha level, that is a higher risk of error, can be the only viable solution when the sample size cannot be increased (Biffl and Halling, 2003; Dunsmore et al. 2003; Laitenberger et al. 2001). In our case, the sample size could not be increased because of the restrictions posed by the course and the constraint to follow a pure volunteer-based participation. We can only acknowledge that the small sample of experimental subjects represents a useful, yet less than ideal circumstance in which we furthered our understanding in the field of CMC applied to requirement engineering, and replications of our study in settings with more resources available are encouraged.

6.4.2 Threats to Internal Validity

Threats to *internal validity* influence the conclusions about a possible causal relationship between the treatment and the outcome of a study. We identify the following rival explanations for our findings.

An *instrumentation effect* occurs when differences in the results may be caused by differences in experimental material. Because we evaluated the interaction between stakeholders who defined the requirements for six different applications, the differences in the application domain and complexity may have acted as confounding factors. However, since the projects had to conform to a number of criteria as indicated in Section 5.1, this threat was partially mitigated. Another possible threat is that we could not eliminate, or limit, the face-to-face interaction of the groups involved in TXT interactions, and which could possibly have affected their performance. We believe that this is not a significant threat because it is only in rare situations that project member interaction is completely limited to text-based

communication and a certain amount of synchronous communication either through audio or video conferences, or face-to-face interactions, occurs even in distributed projects.

6.4.3 Threats to External Validity

External Validity Describes the study representativeness and the ability to generalize the results outside the scope of the study. We identified the following threats to external validity.

Representative Subjects Our students may not be representative of the entire population of software professionals and, similar to any academic laboratory experiment, the ability to generalize results to the industrial practice of RE is limited. However, experiments with students from software engineering courses should not be overlooked. If appropriately trained, students as subjects are acceptable and if the data establishes a trend. Based on the behavioral theory of group performance, Sauer et al. (2000) state that task expertise is the dominant determinant of performance in software engineering activities; they recommend training to increase skills. Since this experiment was part of a specific course on RE, we had a chance to train students on the applicable techniques and the requirements definition process during a semester. The issue with the representativeness of study participants is related to their familiarity with the use of synchronous, text-based communication. Computer science students are very accustomed to text-based interaction. Nevertheless, synchronous, text-based communication tools, such as chat and IM, are increasingly adopted in the workplace, not only in the field of software development, to complement email (Herbsleb et al. 2002).

Representative Process Nonetheless, the requirements definition process in this experiment may not be representative of industrial practice. Yet, unlike many other experiments on media effects in the literature, we did not use generic, puzzle-like tasks that involve idea generation or problem solving. Instead, we designed our experimental tasks for elicitations and negotiations of software requirements for non-toy-sized, realistic systems. Our study required the development and use of specialized knowledge from several different documents during the RE process (e.g., the RFP during the elicitation workshops, the RS 1.0 during the negotiations), as well as techniques learned through the course (e.g., meeting facilitation), to accomplish the tasks. This resulted in a high cognitive load for the study participants and a realistic effort required for accomplishing the experimental tasks.

Similarly, we simulated geographical dispersion. We mitigated the threat of face-to-face interaction during the TXT requirements workshops by distributing the members in each team over three laboratory rooms asserting strict control over student interaction to prevent them from interacting verbally throughout the workshops.

Despite our efforts to counteract or mitigate these threats to external validity, they are inherent to running academic laboratory experiments and can be completely overcome only by conducting replications with people, products, and processes from an industrial context.

6.4.4 Threats to Construct Validity

Construct validity Concerns the degree of accuracy to which the variables defined in the study measure the constructs of interest. We identified the following threats in our study.

Operationalizations Accuracy The constructs of *satisfaction with performance* and *comfort with communication mode* have been adapted from a similar study on media effects (Murthy and Kerr 2000). The several questions used to measure these constructs were defined by taking

into account (1) the communication issues commonly experienced and already acknowledged by previous research in RE (e.g., Al-Rawas and Easterbrook 1996; Gottesdiener 2002); and (2) the issues informally reported by the students. While one could argue arbitrariness in the definition of the scales used to operationalize each construct, we overcome this threat by executing principal component analysis and scale reliability analysis. In particular, to ensure the validity of the constructs, we performed factor analysis with varimax rotation and a cut-off point of .70. Additionally, we performed scale reliability analysis to determine further the construct validity by assessing the extent to which a set of questions measures a single latent variable. For this purpose, we computed the Cronbach's alpha coefficient, which represents the most-widely used index of internal consistency in social sciences (Cronbach 1951). The alpha indexes of the extracted component were .82 and .75 for the scale in the post-elicitation questionnaire and the post-negotiation questionnaire, respectively. Both values are above the threshold of .70, allowing us to affirm scale reliability (Nunnally 1978).

With regard to the construct of *lack of common ground*, because the measures were obtained from the content analysis of the electronic workshop logs, we mitigated the threats to construct validity by using two independent coders and measuring the intercoder agreement. Furthermore, we compared TXT elicitations and negotiations to establish their goodness of fit to facilitate the establishment of common ground. However, depending on the process they followed (see Fig. 5), stakeholders had a different number of prior F2F meetings (i.e., participants in TXT negotiations had already taken part in F2F elicitations) and, thus, had different opportunities to build common ground. Any F2F interaction fosters the establishment of common ground that positively affects future CMC interactions. Although we acknowledge this, we point out that it runs counter to our H3 hypothesis and findings. In fact, the results show that, despite their higher entry level of common ground, participants in TXT negotiations manifested more evidence of lack of common ground than did participants in TXT elicitations. Finally, with respect to the construct of *quality of RS documents*, we identify two limitations. First, although the defects collected through inspections represent an acknowledged standard indicator of document quality, they might not have fully captured quality in terms of completeness. For instance, the RS 2.0 document from Project3, although one of the best quality with only 8 major defects, was also missing the prototypes and mockups of the system, which counted as only one major defect despite its severity. Second, while one can reasonably assess the quality of a (communication) process by measuring the quality of its final output, we must point out that the specification documents were not completely developed during the requirements workshops. Instead, the documents were finalized during follow up tasks, which might have affected the quality of the specifications as well as the communication issues raised during the requirements workshops.

7 Conclusions and Future Work

Our empirical study compared the effectiveness of synchronous text-based communication with face-to-face communication for supporting distributed requirements elicitations and negotiations. Our findings challenge the common belief that face-to-face communication is the most appropriate medium for requirements workshops, and suggest conditions and reasons for preferring text-based communication would be preferred for such interaction. We hope that further work will replicate our empirical study and produce more confidence in these results, as well as test the support of other communication modes positioned between face-to-face and synchronous text-based communication with respect to the richness dimension in our framework (e.g., audio/video-conferencing).

The development of our study hypotheses was guided by a framework of analysis of computer-support for communication in requirements elicitations and negotiations. Based on existing theories of computer-mediated communication and media selection for group work, our theoretical framework considered both the intricate properties of communication media and the specific information richness requirements of requirements elicitation and negotiations tasks. This analysis underlined that the technology fit to requirements tasks needs understanding on two levels, perceived and actual. The properties of rich and lean communication media allowed us to hypothesize about the performance of groups engaged in requirements elicitations and negotiations, both from a subjective and an objective evaluation perspective. We obtained partial support to our hypotheses.

We hope that future research will find this framework useful in testing additional hypotheses about computer-support for the communication-intensive tasks of RE, and in providing the necessary terminology to engage in a debate about the information richness inherent in requirements tasks. Albeit a first attempt—to the best of our knowledge—to analyze systematically computer support for the tasks of requirements elicitations and negotiations, this framework is now accompanied by empirical evidence that makes its further development into a model of CMC support for requirements tasks necessary. Our study identified specific aspects of requirements tasks that may not have been accounted for by previous theories of CMC. For example, F2F was only in some circumstances the preferred medium for requirements tasks, and requirements elicitations on lean media did not necessarily underperform those in F2F elicitations. This suggests that future work should consider, through repeated or replicated empirical studies such as the one described here, the development of models that specify CMC support for requirements tasks more precisely than the hypotheses we drew from existing theories of CMC.

Acknowledgements We thank all the students who took part in the experiment and the RS inspectors. We also thank the anonymous reviewers for their valuable feedback.

Appendix: Experimental Data

Appendix 1 Description of projects and teams used in the empirical investigation

Project		Team	Description
Project 1	UVic Centre for Scholastic Entertainment Edu Game	Gr1 (5 clients) Gr2 (8 developers)	An educational game designed to help students in grades 1 & 2 with Math, English and Problem-Solving skills
Project 2	Equipment and Patient Tracking for St. Peter Hospital.	Gr2 (8 clients) Gr3 (6 developers)	A system to keep track of supplies, equipment and patients of St. Peter Hospital
Project 3	Bus Tracking System	Gr3 (6 clients) Gr4 (6 developers)	A bus tracking system to assist passengers with route planning, time schedule, and connecting busses
Project 4	G4-consulting Groupwork System	Gr4 (6 clients) Gr5 (6 developers)	A collaborative development suite to improved developers collaboration
Project 5	University of Vancouver Island Room Organization System	Gr5 (6 clients) Gr6 (7 developers)	A centralized web based system used to view and book room resources around campus for various events
Project 6	SysCal Shared Calendar	Gr6 (7 clients) Gr1 (5 developers)	A shared calendar for arranging meetings/ schedules of company employees.

Appendix 2. Categories of the coding schema (thematic units) in content analysis

Category		Description
QUESTION	YES-NO	A question that takes a yes/no answer or just a few words (e.g., Q: *"How many beds are available in the hospital overall?"*, A: *"There are 400"*). It may initiate a grounding chain.
	COMPLEX	Any other question not covered by the YES-NO QUESTION category, such as the wh- questions or those that aggregate multiple questions in a single utterance (e.g., *"What information will each employee use to login to the system? Is a key card a consideration?"*). It may initiate a grounding chain.
ANSWER	YES-NO	An answer to a question that takes a few words or a yes-no utterance (e.g., *yes*, *no*, *yep*, *nope*, *y*, *n*, *yeah*, *"yeah, 400"*). Normally appears after a YES-NO QUESTION or CHECK. It may close a grounding chain.
	COMPLEX	Any other answer to a question not covered by the YES-NO ANSWER category (e.g., Q: *"Are you going to do the display public events portion for the final product?"*, A: *"No, they're shown elsewhere"*). It may close a grounding chain.
CHECK	PROVISIONAL	Any utterance that explicitly looks for confirmation of acceptance through provisional, try-marked statements (e.g., *"So we decided for 400 beds, right?"*). It is normally followed by an AGREEMENT or an ANSWER.
	VERBATIM COPY	Any utterance that explicitly gives confirmation of acceptance by verbatim copying a previous utterances (e.g., *"Backup monthly on a tape"*, *"Ok, once a month on tape"*). It is normally followed by an AGREEMENT.
	MISUNDERSTANDING	Any utterance that provides evidence that a previously entered utterance was not accepted (e.g., *"I'm not sure I get the question"*, *"What?"*). It may initiate a grounding chain and is normally followed by a TASK or an ANSWER.
ACKNOWLEDGMENT		Any utterance that explicitly demonstrates that a previously entered utterance has been understood and accepted (e.g., *ok*, *k*, *fine*), but not after a CHECK or QUESTION. It may close a grounding chain.
TASK		Any task-related utterance, presented not in response to a question, which does not express acknowledgement or (dis)agreement (e.g., for providing clarification or extra information).
AGREEMENT		Expresses agreement with a previously entered utterance, but not as an affirmative answer to a question, including smileys (e.g., *yes*, *yep*, *y*, *k*, *yeah*, *ok*, *right*, *I see*, *I agree*). It normally appears after a QUESTION, CHECK, or TASK utterance and may also close a grounding chain.
DISAGREEMENT		Expresses disagreement with a previously entered utterance, but not as a negative answer to a question (e.g., *no*, *nope*, *n*). It may also initiate or continue a grounding chain.
REPAIR		Any fragment entered to repair an error, typically in case of typos (e.g., *"(The system) will use various (calendar formats) depending on the information displayes"*, *"...displayed"*).
OTHER		Off-topic communication, not related to task, such as technical issues, preparation, and social messages. It may include smileys (e.g., *"I got disconnected here"*, *"Sorry, I'm late!"*, *"LOL!"*).

Appendix 3. An example of a grounding chain with categories

Unit	Sender	Content	Category
1	Developer1	How long do patient records remain on the system?	Y/N Question
2	Client1	Forever	Y/N Answer
3	Developer1	Okay	Acknowledgment
4	Client2	No wait	Disagreement
5	Client1	?	Check Misunderstanding
6	Client2	if the patients have checked out for over a month then their information should be archived but still remain accessible if need be	Task
7	Client3	Yeah patient records are permanent	Agreement
8	Developer1	Ok, records archived after one month	Check Verbatim Copy
9	Client3	Yeah	Agreement

References

Al-Rawas A, Easterbrook S (1996) Communication Problems in Requirements Engineering: a Field Study. Proc. 1st Conf. on Professional Awareness in Software Engineering, 47–60

Andrews DC (Mar. 1991) JAD: a crucial dimension for rapid applications development. J Syst Manag 42(3):23–31

Biffl S, Halling M (May 2003) Investigating the defect detection effectiveness and cost benefit of nominal inspection teams. IEEE Trans Softw Eng 29(5):385–397

Birnholtz JP, Finholt TA, Horn DB, Bae SJ (2005) Grounding Needs: achieving common ground via lightweight chat in large, distributed, ad-hoc Groups. Proc. Int'l Conf. on Human Factors in Computing Systems (CHI'05), 21–30, doi:10.1145/1054972.1054976

Calefato F, Lanubile F (2009) Using frameworks to develop a distributed conferencing system: an experience report. Software Pract Ex 39(15):1293–1311. doi:10.1002/spe.937, ISSN: 0038–0644

Campbell CL, Van de Walle B (2003) Asynchronous Requirements Engineering: Enhancing Distributed Software Development. Proc. Int'l Conf. Information Technology: Research and Education (ITRE'03), 133–136, doi:10.1109/ITRE.2003.1270588.

Carmel E, Agarwal R (2001) Tactical approaches for alleviating distance in global software development. IEEE Softw 18(2):22–29. doi:10.1109/52.914734, Mar/Apr 2001

Cheng B, Atlee J (2007) Research directions in requirements engineering. Future of Software Engineering, Minneapolis, MN, USA, 285–303, doi: 10.1109/FOSE.2007.17

Clark HH, Brennan SE (1991) Grounding in Communication. In: Perspectives on Socially Shared Cognition. American Psychological Association, Washington DC, 127–149

Conover WJ (1980) Practical Nonparametric Statistics. Wiley, New York

Coughlan J, Macredie RD (2002) Effective communication in requirements elicitation: a comparison of methodologies. Requir Eng J 7(2):47–60

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16:297–334

Cummins RA, Gullone E (2000) Why we should not use 5-point Likert scales: the case for subjective quality of life measurement. Proc. 2nd Int'l Conf. quality of life in cities, 74–93

Curtis B, Krasner H, Iscoe N (1988) A field study of the software design process for large systems. Commun ACM 31(11):1268–1287

Daft RL, Lengel RH (1986) Organizational information requirements, media richness and structural design. Manag Sci 32(5):554–571

Damian D (2007) Stakeholders in global requirements engineering: lessons learned from practice. IEEE Softw 24(2):21–27, Mar./Apr

Damian D, Zowghi D (2003) Requirements challenges in multi-site software development organizations. Requir Eng J 8(3):149–160

Damian D, Eberlein A, Shaw ML, Gaines BR (2000) Using different communication media in requirements negotiation. IEEE Softw 17(3):28–36, May/Jun. 2000

Damian D, Eberlein A, Shaw ML, Gaines BR (2003) An exploratory study of facilitation in distributed requirements engineering. Requir Eng 8(1):23–41

Damian D, Lanubile F, Mallardo T (Jan. 2008) On the need for mixed media in distributed requirements negotiations. IEEE Trans Softw Eng 34(1):116–132
Dennis AR, Valacich JS (1998) Testing media richness theory in the new media: the effects of cues, feedback, and task equivocality. Inf Syst Res 9(3):156–274
Dennis AR, Fuller RM, Valacich JS (2008) Media, tasks, and communication processes: a theory of media synchronicity. MIS Q 32(3):575–60, Sept
Dunsmore A, Roper M, Wood M (August 2003) The development and evaluation of three diverse techniques for object-oriented code inspection. IEEE Trans Softw Eng 29(8):677–686
Fjermestad J, Hiltz R (Dec. 1998) An assessment of group support systems experimental research: methodology and results. J Manag Inf Syst 15(3):7–149
Garland R (1991) The mid-point on a rating scale: is it desirable? Mark Bull 2:66–70
Gottesdiener E (2002) Requirements by collaboration: Workshops for defining needs. Addison-Wesley Longman
Gunawardena CN, Zittle FJ (1997) Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. Am J Dist Educ 11(3):8–26
Hatcher L, Stepanski EJ (1994) Step-by-Step Approach to Using the SAS System for Univariate and Multivariate Statistics. SAS Inst. Inc., Cary, N.C
Herbsleb JD, Mockus A, Finholt TA, Grinter RE (2001) An empirical study of global software development: distance and speed. Proc. Int'l Conf on Software Engineering (ICSE'01), 81–90, doi:10.1109/ICSE.2001.919083
Herbsleb JD, Atkins DL, Boyer DG, Handel M, Finholt TA (2002) Introducing instant messaging and chat into the workplace. Proc. Int'l Conference on Computer-Human Interaction (CHI '02). Minneapolis, USA, 171–178, doi:10.1145/503376.503408.
Laitenberger O, El Emam K, Harbich TG (May 2001) An internally replicated quasi-experimental comparison of checklist and perspective-based reading of code documents. IEEE Trans Softw Eng 27(5):387–421
Lloyd WJ, Rosson MB, Arthur JD (2002) Effectiveness of elicitation techniques in distributed requirements engineering. Proc. IEEE Int'l Conf. on Requirements Engineering (RE'02), 311–318, doi:10.1109/ICRE.2002.1048544
Macaulay LA (1996) Requirements Engineering. Springer-Verlag Telos .
Maiden N, Manning S, Robertson S, Greenwood J (2004) Integrating creativity workshops into structured requirements processes. Proc. 5th Conf. on designing interactive systems: processes, practices, methods and techniques (DIS'04). Cambridge, MA, USA, 113–122
Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60
McGrath JE (1984) Groups: interaction and Performance. Prentice Hall
McGrath JE (1991) Time, interaction, and performance (TIP): a theory of groups. Small Group Res 22 (2):147–174
McHoul AW, Rapley M (2001) How to analyse talk in institutional settings: a casebook of methods. Continuum International Publishing Group
Montgomery DC (1996) Design and Analysis of Experiments, 4th edn. Wiley, New York
Murthy US, Kerr DS (2000) Task/technology fit and the effectiveness of group support systems: evidence in the context of tasks requiring domain specific knowledge. Proc 33rd Hawaii int'l conf On system sciences (hicss-33) 2:1–10. doi:10.1109/HICSS.2000.926611
Nunnally J (1978) Psychometric Theory. McGraw-Hill, New York
Nuseibeh B, Easterbrook S (2000) Requirements Engineering: a roadmap. Proc. Int'l Conf. on the Future of Software Engineering (ICSE'00), 35–46, doi:10.1145/336512.336523
Olson JS, Olson GM (2000) Distance matters. Hum Comput Interact 15(2/3):139–178
Olson JS, Olson GM, Meader DK (1995) What Mix of Video and Audio is Useful for Remote Real-Time Work? Proc. Int'l Conf. on Human Factors in Computing Systems (CHI '95), 362–368, doi:10.1145/223904.223951
Robert LP, Dennis AR (2005) Paradox of richness: a cognitive model of media choice. IEEE Trans Professional Comm 48(1):10–21. doi:10.1109/TPC.2004.843292, Mar/Apr
Sauer C, Jeffery DR, Land L, Yetton P (2000) The effectiveness of software development technical reviews: a behaviorally motivated program of research. IEEE Trans Softw Eng 26(1):1–14
Seaman CB (1999) Qualitative methods in empirical studies of software engineering. IEEE Trans Softw Eng 25(4):557–572
Short J, Williams E, Christie B (1976) The Social Psychology of Telecommunications. Wiley, New York
Sommerville I, Sawyer P (1997) Requirements Engineering: A Good Practical Guide. Wiley, New York
Stemler S (2001) An overview of content analysis. Practical Assessment, Research and Evaluation 7(17)
Sutton DC (2000) Linguistic problems with requirements and knowledge elicitation. Requir Eng J 5(2):114–124
Taylor SJ, Bogdan R (1984) Introduction to Qualitative Research Methods. Wiley, New York
Ten Have P (1999) Doing conversation analysis: a practical guide. Sage, 1999.
Titscher S, Jenner B, Meyer M (2000) Methods of text and discourse analysis: in search of meaning. Sage

Veinott E, Olson JS, Olson GM, Fu X (1999) Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. Proc. Int'l Conf. Computer-Human Interaction (CHI'99), 302–309, doi:10.1145/302979.303067.

Walther JB (1996) Computer-mediated Communication: impersonal, interpersonal, and hyperpersonal interaction. Commun Res 23(1):3–43

Walz DB, Elam JJ, Curtis B (1993) Inside a software design team: knowledge acquisition, sharing and integration. Commun ACM 36(10):63–77

Wohlin C, Runesson P, Höst M, Ohlsson MC, Regnell B, Wesslén A (2000) Experimentation in Software Engineering. An Introduction, Kluwer Academic Publishers

Zigurs I, Buckland BK (Sep 1998) A theory of task/technology fit and group support systems effectiveness. MIS Q 22(3):313–334. doi:10.2307/249668

Fabio Calefato received the MSc and PhD degrees in computer science from the University of Bari, Italy. Currently, he is associated with the Collaborative Development Group at the University of Bari as a postdoctoral research assistant. His research interests are focused on studying computer-mediated communication and global software engineering.

Daniela Damian is an associate professor in the Department of Computer Science at the University of Victoria and the leader of the Software Engineering Global interAction Laboratory (SEGAL). Her interests includes collaborative software engineering with a special interest in global software development. She serves on the editorial boards of the Journals of Empirical Software Engineering, Requirements Engineering, and Software Process Improvement and Practice, as well as on program committees of conferences including Requirements Engineering, Software Engineering, Global Software Development, and Foundations of Software Engineering. She is a member of the IEEE Computer Society.

Filippo Lanubile is an Associate Professor of computer science at the University of Bari where he leads the Collaborative Development Research Group. His research interests lie in the areas of software engineering and computer supported cooperative work, focusing on social software engineering and distributed software development. He is a recipient of a NASA Group Achievement Award (1996, as a research associate at the University of Maryland), an IBM Eclipse Innovation Award (2006), an IBM Faculty Award (2008) and the Software Engineering Innovation Foundation Award from Microsoft Research (2011). Filippo Lanubile has been the Program Chair of the 2008 edition of the IEEE Int. Conf. on Global Software Engineering (ICGSE) and will be the General Chair of the ICGSE 2013 conference.