Learning from virtual interaction:
A review of research on online synchronous groups

Anindito Aditomo and Peter Reimann
CoCo Research Centre, University of Sydney, NSW 2006, Australia
aadi4954@usyd.edu.au; p.reimann@edfac.usyd.edu.au

Abstract: Although in general collaborative learning is effective, it is clear that this is not always the case. To explain this, researchers have been suggested to investigate the interaction process occurring in the course of collaboration. Research on face-to-face (FTF) groups have provided clues as to what types of interaction are productive for learning, both at the individual and group level. However, the extent to which these findings apply to online groups is not yet clear. This paper reports a conceptual systematic review of recent studies of online synchronous learning groups. There is little evidence that the types of online interaction deemed favorable are actually associated with individual conceptual learning. These findings challenge the implicit assumption held by many educational technology designers. Implications for future research are discussed.

Introduction

Two heads, more often than not, are better than one. This commonsense wisdom apparently applies also to learning: learning in groups is generally more effective than alone (Cohen, 1994; Johnson & Johnson, 2004). However, research has also found that collaboration does not always lead to better learning outcomes (for a recent metaanalysis, see Springer, Stanne, & Donovan, 1999). To explain this, researchers have been suggested to look into the collaboration process (Dillenbourg, Baker, Blayc, & O'Malley, 1995). Indeed, studies have found that certain forms of interaction process are linked to learning. For example, giving explanations during peer-directed mathematics study groups has been found to be related with subsequent individual achievement (Webb, 1982, 1991). Similarly, interpretive talk, but not descriptive talk, between dyads working to solve a programming problem has been found to be related with group performance and individual understanding (Teasley, 1995).

These and other findings (Barron, 2000a, 2000b, 2003; Chan, 2001; Chan, Burtin, & Bereiter, 1997; Kneser & Ploetzner, 2001; Oshima, Scardamalia, & Bereiter, 1996) demonstrate that certain forms of interaction are associated with individual learning and group performance. However, most of these findings come from studies of face-to-face (FTF) groups. Will the same links be found in online synchronous learning groups? What forms of interaction are associated with individual learning and group performance of online groups? These questions are important because many online learning environment and cognitive tools are designed with an eye to facilitating certain forms of interaction, which are assumed to bear learning benefits.

But why should we suspect that interaction of online groups would be any different from FTF groups? We know that different media have different constraints and affordances for communication. Online, text-based communication affords more persistence of information, meaning that previous utterances do not "evaporate" immediately, as they are recorded in the chat environment. On the other hand, it is limited in terms of emotional expressions, deictic gestures, spontaneous response, and eye gaze, which are subtle but important in achieving "common ground" (Clark & Brennan, 1991). Because of these differences, achieving a grounding criterion sufficient for learning to occur would entail a different process for online and FTF groups.

Furthermore, detailed analysis of groups solving complex conceptual problems shows that successful collaboration is based on the co-construction of a joint problem space (Roschelle, 1995). More recently, another study has proposed that collaboration involves two spaces: a *content space* (which is more cognitive and associated with the problem to be solved), and a *relational space* (which more to do with affective and social aspects of interaction, such as identity and conflict) (Barron, 2003, p. 310). For groups to maintain a joint attention that is productive for individual learning and group performance, these two spaces must be coordinated well. How this complex coordination is achieved, once again, would differ with regards to the medium of communication. Thus, it is quite reasonable to examine how interaction and learning links specifically in computer-mediated groups.

Objectives, scope, and approach

This review will focus on studies of online, synchronous learning groups, and thus will extend previous reviews of FTF collaborative or cooperative learning (Cohen, 1994; Webb, 1982, 1991). Unlike most previous reviews, however, the purpose here is not to assess the effectiveness of online collaboration. Instead, this review seeks to examine what kinds of online interaction processes are associated with learning outcomes, both at the individual and group level (group problem solving performance). This was done by systematically searching empirical reports from nine representative journals relevant to the current purpose[1]. In addition, articles from the 2005 CSCL Conference were also systematically searched. The search used several combinations of keywords ("interaction", "collaborative learning", and "computer").

The retrieved articles were then selected based on the following criteria: published recently (from 2000 until mid-2006), report data on both interaction processes and learning outcomes, and investigated online groups communicating synchronously. These limitations excluded many articles that report interaction data but not learning outcome (and vice versa), and studies of groups collaborating asynchronously (e.g., using emails or wikis). This review is admittedly selective and not comprehensive, as the aim is more towards finding conceptual insights than summarizing empirical results (as in meta-analyses). However, the aforementioned criteria were used to limit possible bias or subjectivity on the inclusion of studies from the search.

Description of studies

The selection process resulted in 12 articles, covering 606 groups (1161 individuals). From these studies, 14 separate results can be examined (one article reported 2 experiments, and another article reported 2 different measures of individual learning). A brief description of these studies is given below:

Participants and tasks

Participants were school-aged or university students who worked in pairs or groups of three (triads). The collaboration was relatively short-term, with most studies (9 of 12) using single-session meetings of 40 to 120 minutes. The remaining studies used multiple sessions of 3 to 6 meetings. All but one study used tasks that were domain-related, such as tasks in physics (fluid dynamics), biology (heredity, food chain), psychology (clinical case study and attribution theory), and historical inquiry. One exception, which is a more general problem solving task, was solving a murder case. Some tasks were ill-structured (e.g., writing argumentative essays and constructing a clinical diagnosis), while others are somewhat more constraining (e.g., constructing a concept map using a set of given concepts, or answering multiple choice questions).

Types of learning scaffold

All but one study used scaffolds that were built into the software environment as cognitive tools. One exception is Saab, van Joolingen, & van Hout-Wolters' (2006) study, which used a collaboration script (the RIDE rule) delivered as an instruction prior to the interaction. The rest of the studies used cognitive tools, which can be categorized into script-prompts and external representation tools. Scripts are prompts built into the learning environment that facilitate certain actions, be it epistemic acts (e.g., finding data to ground certain claims) or social acts (e.g., taking a certain role in the interaction). External representation tools vary in their degree of constraint, ranging from very generic and "loose" (such as a virtual whiteboard) to more constraining (such as concept maps and dynamic models).

Approaches to coding interaction data

The coding used by most of the studies reviewed here can be categorized into three approaches. *The first*, simplest way of "coding" interaction is merely counting the number of utterance or message, disregarding the content or meaning of the message. *The second* approach differentiates the content of single utterances

---

[1] These are *Computers and Education, Computers in Human Behavior, Cognition and Instruction, Educational Technology Research and Development, Instructional Science, Journal of Computer Assisted Learning, Journal of Educational Computer Research, Learning and Instruction,* and *The Journal of the Learning Sciences*

using a certain categorization. One example is Chiu's (2004) coding, which counted the number of "knowledge-related" utterances produced by learners during a collaborative concept mapping session. A rather more refined example is the coding of "task acts" in van Drie, van Boxter, Jaspers, & Kanselaar's (2005) study, which categorized utterances as on-task, procedural, technical, social talk, or greetings. *The third* approach parses interaction data at the episode level, which is a series of single moves or utterances. Included in this approach is van Drie et al.'s (2005) coding of episodes into domain-specific reasoning, elaborated reasoning, and co-constructed reasoning. It should be noted that these coding approaches were sometimes used in combination by one study. One study (Dillenbourg & Traum, 2006) used a somewhat different approach. Rather than analyzing utterances and episodes, Dillenbourg and Traum looked into degree of grounding (see Clark & Brennan, 1991), as indicated by how often participants explicitly acknowledge that they have understood their partner's move. To summarize, we can differentiate between coding of interaction at the individual/solo action level (single utterances) as opposed to coding at the joint action level (episodes of utterances and degree of grounding).

Assessment of learning outcome

Learning can be assessed at the individual and group level. Assessment of individual learning included declarative knowledge, problem-solving transfer, and reproduction of group problem-solving behavior (e.g., constructing a concept map using the same set of concept and relations). With regard to group performance, we can distinguish between assessment of problem representation produced during collaboration (such as concept map or dynamic model) and problem solution. The problem solution category can be distinguished further into responses to structured tests and construction of artifacts (such as essays).

Results

Results of this review will be discussed separately for learning outcomes measured at the individual level and those measured at the group level (group performance in problem solving).

Interaction and individual learning

There seem to be little evidence that online interaction is actually related with individual learning outcomes (see Table 1). Two studies of groups learning (Makitalo, Weinberger, Hakkinen, Jarvela, & Fischer, 2005; Weinberger, Ertl, Fischer, & Mandl, 2005) found that producing more messages or words during interaction was not associated with higher ability in a near transfer posttest (applying conceptual understanding of Weiner's attribution theory to a new case). Further unfavorable evidence comes from Zumbach, Schonemann, and Reimann (2005), who investigated dyads communicating via a chat to construct a clinical diagnosis of a psychological disorder patient. These authors analyzed the interaction process at the episodic level, looking at action-response sequences defined as collaborative events. No correlation was found between the frequency of collaborative events and individual gain in declarative knowledge test on the topic (depression and anorexia-nervosa), or quality of the clinical diagnosis.

These results are somewhat unsurprising, as the study authors used only a measure of frequency. The first two studies (Makitalo et al., 2005; Weinberger et al., 2005) simply counted the number of words or messages. Zumbach, Schonemann, and Reimann (2005) also did not differentiate between types of collaborative events (although coded interaction at the episode level) and thus lump together short sequences (e.g., simply agreeing to a proposal put forward the partner) with more elaborated sequences (e.g., challenging a proposal and then elaborating reasons for a counter-proposal). However, it is important to note that dyads in one of the experimental conditions did not engage in any collaborative event, but nevertheless achieved high individual learning outcome. This means that collaborative interaction (whether short or elaborated) as defined in this study does not necessary relate to quality of individual learning.

Table 1: Evidence of link between interaction and individual learning

Author (year)	Topic/domain	Coding of interaction	Assessment of learning outcome	Supporting Evidence
(Weinberger et al., 2005)	Theory of genotype environment effect	Utterance content	Declarative knowledge	Positive

(Chiu, 2004)	Heredity, food chain, & atmosphere	Utterance content	Individually reproducing a concept map previously performed collaboratively)	Positive
(Weinberger, Fischer, & Stegmann, 2005)	Attribution theory	Utterance content	Constructing arguments	Positive
			Problem-solving transfer	Negative
(Makitalo et al., 2005)	Attribution theory	Frequency level	Problem-solving transfer	Negative
(Weinberger et al., 2005)	Attribution theory	Frequency level	Problem-solving transfer	Negative
(Zumbach, Schonemann, & Reimann, 2005)	Clinical psychology	Episode content	Declarative knowledge	Negative
			Problem-solving transfer (writing clinical diagnosis)	Negative
(van Drie et al., 2005)	Historical inquiry	Utterance content	Declarative knowledge	Negative
(Saab, van Joolingen, & van Hout-wolters, 2006)	Physics	Utterance content	Declarative knowledge	Negative
			Problem solving transfer	Negative

A further question is whether certain types of online interaction (for example, elaborated or knowledge related episodes) are more associated with individual learning. Results from van Drie et al.'s (2005) study of dyads performing a historical inquiry task facilitated by several different representational forms (diagram, matrix, or list) can be used to address this question. In coding the interaction process, van Drie et al. identified episodes when dyads engaged in elaborated explanations. However, although the matrix and control groups produced more historical reasoning, co-construction, and co-elaboration episodes, they did not perform better in a subsequent individual test on historical reasoning. Similarly, in studying pairs solving problems in a physics microworld, Saab, van Joolingen, and van Hout-Wolters (2006) found that groups who engaged in more "communicative activities", "discovery transformative activities" (e.g., describing and recognizing relations and drawing conclusions), and "regulative transformative activities" did not perform better in declarative nor near-transfer knowledge tests.

Further unfavorable evidence comes from Weinberger et al. (2005), who studied triads communicating via a message board to analyze cases using attribution theory. In this study, groups provided with argumentation scripts produced more counter-arguments and twice as many grounded claims in their messages. This indicates that these group members engaged in deeper cognitive processes related to the learning material. Despite this, no differences were found in individual members' performance in analyzing a new case using attribution theory. However, it is interesting to note the groups using argumentation scripts did outperform other groups in a posttest measuring ability to construct arguments. Thus it seems that although the interaction quality has little influence on domain knowledge, it does have a positive influence on a more general epistemic skill (constructing arguments) that was performed or practiced during the collaboration.

Evidence of this last point can also be found in Chiu's (2004) study of 6th grade student triads constructing concept maps of several domain-knowledge. Students who produced more knowledge-related utterances during their collaboration had better ability to reconstruct the concept map (of the same set of concepts and relations) two weeks after the experiment. Again, this is evidence that relevant interaction during collaboration is related to an epistemic skill (which is, in this case, the epistemic skill of constructing a semantic network) performed during the collaboration.

The only evidence that elaborating knowledge-related material during collaboration is related to acquisition of declarative knowledge comes from Weinberger et al.'s (2005) study of pairs studying a theory of genotype environment effect. Communicating via a video conference channel, students who engaged in more in theory elaborations also gained higher score on a cued-recall test. It is important to note that in this study, the context was a peer-tutorial interaction, which has long been known to be an effective pedagogical approach. Elaborations within this context requires more cognitive processing, as they are performed explicitly or deliberately (explaining and elaborating *is* the group task).

In summary, although certain types of interaction do relate to better ability in performing behavior previously practiced in groups (constructing concept maps and arguments), there is little evidence that they lead to improve declarative or conceptual knowledge. The only exception comes from a study that uses a peer-tutorial scenario and video-conference (i.e. not chat-based).

Interaction and group performance

The studies reviewed provide some evidence for a link between interaction and learning at the group level (measured by a group's performance to produce a solution to a problem, which can be an essay, concept maps, or responses to a set of questions). Some of the studies summarized in Table 2 use multiple indicators for group learning outcome (e.g., essay quality can be measured from its organization, argumentation quality, and audience focus), and thus there can be "mixed" evidence in a single study. As will be discussed, there are several interesting patterns of relationship, again depending on how we assess learning.

Table 2: Evidence of link between interaction and group performance.

Author (year)	Topic/domain	Coding of interaction	Assessment of Learning outcome	Evidence
(van Drie et al., 2005)	Historical inquiry	Utterance & episode content	Essay quality	Mixed
(Erkens, Jaspers, Prangsma, & Kanselaar, 2005)	Organ donation	Episode content	Essay quality	Mixed
(Dillenbourg & Traum, 2006)	Murder case	Degree of grounding	Identifying correct suspect Task management efficiency	Negative Positive
(Chiu, 2004)	Heredity, food chain, & atmosphere	Utterance content	Concept map quality	Positive
(Chiu, Huang, & Chang, 2000)	Computers	Utterance content	Concept map quality	Positive
(Manlove, Lazonder, & de Jong, 2006)	Physics (fluid dynamics)	Episode content	Dynamic model quality	Positive
(Saab, van Joolingen, & van Hout-wolters, 2006)	Physics	Utterance content	Tests of domain-knowledge	Positive
(Saab & Joolingen, 2005)	Physics	Utterance content	Tests of domain-knowledge	Positive

In terms of group learning outcome, we can distinguish between performances in: (1) solving a general logical problem (e.g. a murder case), (2) constructing a textual/linguistic artifact (e.g. argumentative essay), (3) solving a structured domain-specific task (e.g. physics problem), and (4) constructing a graphical external representation (e.g. concept maps and dynamic models). Interestingly, certain types of interaction have a positive effect on the latter two indicators of performance.

In Saab, van Joolingen, & van Houtwolters (2006), performance in solving physics problems is correlated with hypothesis formulation, informative acts, proposing answers, and collecting data in the microworld. Saab and van Joolingen's (2005) earlier study also found similar results. Moreover, these correlations were at the moderate level (0.51 to 0.62). However, they were found only in the experimental groups (not in the control groups), a point that will be returned to later.

Evidence that certain types of interaction influence the quality of graphical representations produced come from three studies. Chiu, Huang, and Chang (2000) found high correlations between on-task interactions with quality of concept map. Chiu's (2004) more recent study also lead to the finding that groups who engaged in more knowledge-related talk also produced better concept maps. In a similar vein, Manlove, Lazonder, and de Jong (2006) found that groups who produced more cognitive episodes in their interaction also constructed better models of a water tank. The correlations between number of cognitive episodes and quality of dynamic model were 0.39 (for experimental groups) and 0.64 (for control groups).

High-quality interaction (e.g. elaboration) does not seem to relate much to group performance in constructing essays (Erkens et al., 2005) or solving a murder mystery (Dillenbourg & Traum, 2006). Both of these tasks are, in a sense, more ill-defined than previously considered tasks. Essay tasks have ill-defined end states (more so than concept maps or dynamic models), whereas murder cases have ill-defined process. However, Erkens did find several weak correlations between coordination processes during collaboration and essay quality. van Drie et al. also found a correlation between co-elaborated episodes with essay quality (r=.66), but only for groups using matrix during their collaboration.

Taken together, these results suggest that certain types of interaction are not directly related to problem solving performance. They are more related intermediate artifacts, such as a group's graphical representation of a problem or a group's efficiency in planning how to go about in an ill-structured task (Dillenbourg & Traum, 2006). However, producing better external representation of a problem, or planning more efficiently, does not guarantee a successful solution.

Discussion and implications

Many CSCL researchers design instructional supports to structure online collaboration with the aim to facilitate certain kinds of interaction. The implicit assumption is that certain kinds (or features) of interaction bear learning benefits. However, the studies reviewed here provide limited support for this assumption. At the individual level, interactions that have been found to be beneficial in FTF groups (e.g. elaborating explanations, co-constructing arguments, and generating grounded claims) do not seem to be related with increase in declarative knowledge or ability to solve problems by applying the conceptual knowledge learnt during collaboration. At the group level, the picture is rather mixed. Certain types of utterance exchanges (episodes) seem to be related to group performance. However, this applies more for intermediate indicators (such as quality of external representation of the problem), but less for end-product indicators (such as essays, which are not representations of the problem, but an end-product of collaboration). Although no conclusive explanations can be offered for these findings, we will try to discuss possible explanations and directions for further research by way of comparison with several published studies of FTF collaboration:

First, many studies of FTF learning groups have found that verbal interaction predicts individual learning gains, even after controlling prior knowledge. Webb (1991), for example, reviewed studies of peer-directed groups instructed to work with mathematical problems. One consistent finding is the positive correlations (controlling for prior math ability) between the frequency of giving elaborated explanations and subsequent individual test. This is in contrast to the relative independence between group processes and individual learning gains found by studies of online groups reviewed here. One possible explanation might be methodological. In most of the studies reviewed here, only one study (Zumbach, Schonemann, & Reimann, 2005) explicitly tested the correlation between measures of process and measures of individual learning outcome. In other studies, this relationship must be inferred from looking at whether groups scoring higher on certain types of interaction also performed better on individual learning. However, this is problematic because these studies aggregated interaction process measures at the group level. For example, an elaborate utterance from one member of a dyad would contribute not only to this particular member, but instead to the dyad's score (the non-contributing member would also be given a score based on his/her partner's contribution). Thus, one direction for future research would be to obtain measures of individual contributions during online collaboration and relate these to individual learning outcomes.

Second, there are several indications that the common coding scheme used to analyze online interaction failed to distinguish utterances reflecting different levels of cognitive activity. This is evident in at least two studies (Saab, van Joolingen, & van Hout-wolters, 2006; van Drie et al., 2005). Saab et al. found that, during collaboration in a physics microworld, utterances reflecting giving information, formulating hypothesis, proposing answers, and collecting data were positively correlated (.51 to .61) with groups' problem-solving performance. However, this was true only for the experimental groups (who were supported with a collaboration scaffold), but not for the control groups. Similarly, van Drie et al. found that the number of co-elaborated episodes a group produced is correlated with its argumentative essay quality ($r = .66$), but this applies only for groups using matrices when constructing their arguments. These indicate that similarly coded utterance or episodes bear different learning consequences, meaning that the coding scheme is somewhat limited in capturing the cognitive value underlying the observable interaction.

One possible explanation is that the "code and count" approach to analyzing group process used in the reviewed studies cannot capture the more emergent properties of the joint discourse. For example, a code and count approach cannot distinguish between two groups producing similar frequencies of a certain type of utterance (elaborated explanation, for example), but embedded in different overall interaction patterns. Several studies of FTF groups indicated that the overall interaction pattern is important. For example, Mercer (1996) noted that groups similarly engaged in a collaborative task can have disputational, cumulative, or exploratory talk patterns (with the first pattern being the least productive for learning). More telling is Barron's (2003) finding from her study of triads collaborating to solve a complex mathematical problem anchored in a video-based story. One of Barron's findings is that groups differed in performance not because none of the members could generate correct solution to the problem. Performance difference was more related to at which point in time during the conversation those correct solutions were proposed. In successful groups, solutions were more frequently proposed in relation with the content of the preceding discussion. This can be interpreted as indicating the importance of sequential structure of collaborative interactions, and call for a sequential approach in analyzing them.

Another way to capture the joint property of group discourse is to use Bereiter's (2002) distinction between problem-centered and referent-centered approaches to knowledge. Using these concepts, we can distinguish between groups who discuss new information as problems to be explained using concepts as tools, and groups who engage in similar discussions but implicitly adopt the goal of accumulating knowledge (which is assumed to reflect reality). Several studies of FTF groups have found that this distinction in how learners approach knowledge corresponds to individual learning and group performance (Chan, 2001; Chan, Burtin, & Bereiter, 1997; Oshima, Scardamalia, & Bereiter, 1996). Whether this distinction is useful for analyzing online collaboration would be another direction for future work.

Third, in contrast to findings from FTF groups, none of the studies reviewed here provide evidence to link interaction through text-based chat with individual mastery of declarative knowledge. A tentative conclusion is that it is difficult to engage in deep cognitive processing in text-based online interaction, especially during short-term collaboration among people who are new to each other. However, (as pointed out by an anonymous reviewer) this will have to be investigated by comparing FTF and online groups using identical materials and measurements. In addition, forms of online interaction do not appear to be related to group performance in joint problem solving. Groups who engage in more meaningful discourse do not necessarily perform better. One possible explanation is that during the collaborative problem solving, groups who did not produce much meaningful discourse achieved similar performance by following their most competent member. This is most evident in one study (Saab, van Joolingen, & van Hout-wolters, 2006), which found a strong correlation between communication asymmetry and group performance among the control groups.

Fourth, there is evidence that text-based online interaction is related with other learning outcomes beside declarative knowledge. In Weinberger, Fischer, and Stegmann's (2005) study, members of groups who produced more grounded claims were better able to construct arguments, which is a valuable epistemic practice. It would be interesting for future research to investigate whether online interaction could help individuals learn other forms of epistemic practices. For example, researchers using concept maps could assess whether participation in collaborative concept mapping helps to enhance an individual's ability to extract core concepts from new texts, or to formulate better guiding questions when comprehending new texts. This is arguably a more meaningful learning outcome than the ability to reproduce the same concept maps from the same set of concepts, as in Chiu's (2004) study. Furthermore, assessment of epistemic practices widens the current focus on domain-knowledge mastery.

References

Barron, B. (2000a). Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences, 9*(4), 403-436.

Barron, B. (2000b). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92*(2), 391-398.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307-359.

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, N.J.: Erlbaum Associates.

Chan, C. K. K. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science, 29*, 443-479.

Chan, C. K. K., Burtin, J., & Bereiter, C. (1997). Knowledge Building as a Mediator of Conflict in Conceptual Change. *Cognition and Instruction, 15*(1), 1-40.

Chiu, C. H. (2004). Evaluating system-based strategies for managing conflict in collaborative concept mapping. *Journal of Computer Assisted Learning, 20*, 124-132.

Chiu, C. H., Huang, C. C., & Chang, W. T. (2000). The evaluation and influence of interaction in networked supported collaborative concept mapping. *Computers & Education, 34*(17-25).

Clark, H. H., & Brennan. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*. Washington: American Psychological Association.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

Dillenbourg, P., Baker, M., Blayc, A., & O'Malley, C. (1995). The evolution of research on collaborative learning. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189-211). UK: Pergamon.

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences, 15*(1), 121-151.

Erkens, G., Jaspers, J., Prangsma, M., & Kanselaar, G. (2005). Coordination processes in computer supported collaborative learning. *Computers in Human Behavior, 21*, 463-486.

Johnson, D. W., & Johnson, R. T. (2004). Cooperation and the use of technology. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology* (pp. 785-810). Mahwah, NJ: Lawrence Erlbaum.

Kneser, C., & Ploetzner, R. (2001). Collaboration on the basis of complementary domain knowledge: Observed dialogue structures and their relations to learning success. *Learning and Instruction, 11*, 53-83.

Makitalo, K., Weinberger, A., Hakkinen, P., Jarvela, S., & Fischer, F. (2005). Epistemic cooperation scripts in online learning environments: Fostering learning by reducing uncertainty in discourse. *Computers in Human Behavior, 21*, 603-622.

Manlove, S., Lazonder, A. W., & de Jong, T. (2006). Regulative support for collaborative scientific inquiry learning. *Journal of Computer Assisted Learning, 22*, 87-98.

Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction, 6*(4), 359-377.

Oshima, J., Scardamalia, M., & Bereiter, C. (1996). Collaborative learning processes associated with high and low conceptual progress. *Instructional Science, 24*, 125-155.

Roschelle, J. (1995). Learning by collaborating: Convergent conceptual change. In T. Koschmann (Ed.), *CSCL: Theory and Practice of an Emergent Paradigm*. Mahwah: Erlbaum Associates.

Saab, N., & Joolingen, W. v. (2005). *Supporting collaborative discovery learning by presenting a tool.* Paper presented at the CSCL???, Tai Pei, Taiwan.

Saab, N., van Joolingen, W. R., & van Hout-wolters, B. H. A. M. (2006). Supporting communication in a collaborative discovery learning environment: The effect of instruction. *Instructional Science*.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21-51.

Teasley, S. D. (1995). The role of talk in children's peer collaboration. *Developmental Psychology, 31*(2), 207-220.

van Drie, J., van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*, 575–602.

Webb, N. M. (1982). Student interaction and learning in small groups. *Review of Educational Research, 52*(3), 421-445.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*(5), 366-389.

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*, 1-30.

Weinberger, A., Fischer, F., & Stegmann, K. (2005). *Computer-supported collaborative learning in higher education: Scripts for argumentative knowledge construction in distributed groups*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Zumbach, J., Schonemann, J., & Reimann, P. (2005). *Analyzing and supporting collaboration in cooperative computer-mediated communication.* Paper presented at the Computer Supported Collaborative Learning 2005: The Next 10 Years!, Tai Pei, Taiwan.

Appendix: Description of empirical studies reviewed

Author (year)	Task context	Participant & interaction time	Learning scaffold	Coding of collaborative interaction	Evidence of link between interaction and learning
(Weinberger et al., 2005)	Peer tutorial via video conference with the goal of understanding a theory of genotype environment effect.	43 pairs (1st year university students). Single session.	Epistemic script (e.g., prompting learners to provide evidence, etc.) & social script (prompting role taking as tutor/tutee).	**Utterance level**: Elaborated utterances, classified into theory elaborations, empirical evidence elaborations, and personal elaborations.	**Individual (positive)**: Dyads provided with social scripts produced more theory elaborations and also gained higher on individual knowledge test (cued recall type).
(Chiu, 2004)	Collaboratively build concept maps from given set of concepts & relations within 3 topics (heredity, food chain, & atmosphere).	32 triads (5th & 6th grade students). Single session of 140 minutes.	Social script that constrain or specify members' role in the concept mapping activity (assigning, rotating, negotiating, and open role).	**Utterance level**: Knowledge-related dialogues.	**Individual (positive)**: The "assign" group produced more knowledge-related talk than all other groups, and performed better in individual knowledge test. **Group (mixed)**: Although the "assign" groups produce more knowledge-related talk, they did not outperform groups in the "give" and "open" conditions. Nevertheless, the "assign" groups outperform the "rotate" groups.
(Weinberger, Fischer, & Stegmann, 2005)	Analyzing 3 cases using attribution theory (e.g., explaining the attribution of a boy about his difficulty with mathematics.	40 triads (1st year university students). Single session of 80 minutes.	Two kinds of argumentation scripts: one designed to foster the construction of single arguments (a text-box for "claims", linked to a text-box for "warrants" and "qualifiers"), and the other for constructing argumentation sequences (much like a dialectical process, going from arguments, counter-arguments, integration, etc.)	**Utterance level**: Single messages were classified based on two dimensions: argumentative and epistemic. In the argumentative dimension, grounded claims (as opposed to simple claims) & counter-arguments were counted. In the epistemic dimension, number application of prior knowledge & new knowledge were counted.	**Individual (mixed)**: Groups given the Single Argument Script produced more counter-arguments & twice more grounded claims than control groups. This is associated with a higher score in a posttest measuring ability to construct single arguments. The Argument Sequence Script groups also produced more counter-arguments, & more instances of prior knowledge application compared to control groups. This is associated with higher posttest score on construction of argumentative sequences. However, no differences were found regarding posttest performance on domain-knowledge.
(Makitalo et al., 2005)	Analyzing 3 cases using attribution theory (e.g., explaining the attribution of a boy about his difficulty with mathematics.	16 triads (university students). Single session of 80 minutes.	Epistemic script designed to guide the case analysis by giving prompts as to what kinds of information to look for.	**Frequency level:** Simply the amount of discourse or number of words produced during the interaction.	**Individual (negative)**: Groups provided with epistemic scripts produced more discourse, but achieved lower in their individual ability to apply attribution theory into a new case (in the posttest).
(Weinberger et al., 2005)	Analyzing 3 cases using attribution theory (e.g., explaining the attribution of a boy about his difficulty with mathematics.	32 triads (1st year university students). Single session of 80 minutes	Epistemic script (see above), and social script, designed to foster collaboration by prompting role taking in the discussion.	**Frequency level**: Simple the number of messages exchanged during the interaction.	**Individual (negative)**: Groups with social scripts produced fewer messages but achieved higher in their individual knowledge test (applying attribution theory to a new case).
(Zumbach, Schonemann, & Reimann, 2005)	Constructing a clinical diagnosis of a patient with psychological disorder.	20 pairs (university students). Single session of 90 minutes.	Feedback as reinforcement of collaboration between partners.	**Episode level**: Certain action-response sequences were coded as collaborative events, which covers relatively short or simple action-response sequence, to more elaborated or complex sequences. Thus, sequences coded as collaborative events were not of equal "quality".	**Individual (negative)**. Number of collaborative events did not correlate with gain in individual knowledge test, nor with quality of clinical diagnosis.
(van Drie et al., 2005)	Historical inquiry task to argue whether or not the changes of Dutch youth in the 1960s were revolutionary, followed by an essay writing task (1000 words).	65 pairs (pre-university students). 6 lessons of 50 minutes.	Different representational forms (diagrams, matrix, and list) designed to facilitate construction of argumentation.	**Utterance level**: Chat was coded based on content, into on-task (including historical reasoning category), procedural, technical, social, and greetings. **Episode level**: Historical reasoning utterances were further coded into episodes, grouped into: domain-specific reasoning, elaboration, and co-construction. The last two were	**Individual (largely negative)**: Although the groups using matrix and control groups produced more historical reasoning, elaboration, co-construction, and co-elaboration (compared to diagram & list groups), they did not performed better in individual post-test of historical reasoning test. **Group (mixed):** The matrix groups did not produce better essay quality. However, there was one significant correlation, which is between number of co-elaborated episodes with essay quality (r=.66), but this was observed only in the matrix group, and not in any other groups.

				combined into a measure of "co-elaboration" episodes.	
(Saab, van Joolingen, & van Hout-wolters, 2006)	Discovering physics laws behind a computer simulation (*Collusions*). Pairs are presented with assignments which require them to experiment or collect data in the microworld.	29 pairs (secondary school students). Single session of 90 minutes.	Instruction on effective collaboration or communication using the RIDE rule (Respect, Intelligent collaboration, Deciding together, and Encouraging), which is also embedded in the online learning environment (pop-up windows prompting certain actions).	**Utterance level**: Utterances and actions categorized into communicative activities (consists of informative, argumentative, elicitative, responsive, directive, and off-task), discovery transformative activities (consists of orientation, generating hypothesis, testing hypothesis, and concluding), and discovery regulative activities (consists of orientation, planning, evaluation, and monitoring).	**Individual (negative)**: Even though the experimental groups engaged in more communicative activities (in terms of deciding together and encouraging acts), discovery transformative activities (in terms of describing & recognizing relations and concluding acts), and discovery regulative activities, they did not perform better in a declarative nor "what-if" knowledge posttests. **Group (mixed)**: Within the experimental group, score in the assignments during collaboration were positively correlated with informative acts (r=.51), formulating hypothesis acts (r= .61), proposing an answer (r=.55), collecting data (r=.55), while correlating negatively with off-task technical acts (r= - .61) and describing & recognizing relations (r=. – 58). In the control groups, score of assignments correlated positively with asymmetry in communication (r=.61), indicating that control groups performed as good as experimental groups because they "followed the leader". This further indicates that group scores is not a good measure of effectiveness of collaboration, because it is highly influenced by the free-rider phenomenon.
(Erkens et al., 2005)	Analyzing information sources to construct an argumentative essay (600 – 1000 words) about organ donation issue.	145 pairs (high school students). 4 – 6 lessons	TC3 (Text Composer, Computer Supported and Collaborative): includes a database of information sources, a private notepad, and a chat facility, plus a certain type of planning tool for writing (argumentation diagram for content generation, and outline for content linearization).	**Episode level:** Chat protocols were coded in 2 broad categories: content-related and writing strategies (Task Acts) and communicative-coordination process (Coordination Process). Task Acts are parsed into planning, executing, and non-task. Coordination Process consists of focusing, checking, & argumentation processes.	**Group (mixed)**: Some weak correlations were found: (a) focusing processes correlates with textual structure of essay (r=.14) and with overall argumentation quality (r=.12), and (b) argumentation processes correlates with overall argumentation quality (r=.13). However, several unexplained correlations were found, particularly some negative correlations between planning acts with text quality (in the control groups).
(Dillenbourg & Traum, 2006)	Solving a murder case: finding the killer from a number of suspects by inferring from clues (e.g., motives, opportunity, etc.).	18 pairs (mostly postgraduate students). Single meeting of 82 – 182 minutes.	Shared whiteboard and MOO environment (text-based virtual reality, in this case includes a virtual map of a hotel, several characters, and objects).	**Degree of grounding**, as indicated by how often pairs explicitly acknowledged their understanding of their partners verbal or other moves.	**Group (mixed)**: The degree of grounding in groups who succeeded to find the murderer did not exceed groups who didn't succeed. However, there is a difference in task management efficiency: the more pairs acknowledged understanding (the more grounding), the more efficient they are in managing or planning the problem solving process.
(Saab & Joolingen, 2005)	Discovering laws of physics by experimenting in a micro-world (to answer open-ended and multiple choice questions).	25 pairs (10th grade students) Single session of 90 minutes.	A tool facilitating hypothesis construction by presenting elements such as variables, plus an evidence palette supporting externalization & evaluation of reasoning.	**Utterance level**: Classified into communicative acts, discovery transformative acts, and discovery regulative acts.	**Group (positive)**: Dyads' performance in solving assignments in the microworld is correlated with the following acts: deciding together (r=.62), transformative (r=.55), & regulative (r=.53). These correlations were found only in the experimental groups using the hypothesis tool.
(Manlove, Lazonder, & de Jong, 2006)	Building a dynamic model of a water tank system (a case in fluid mechanics).	19 triads (high school students). 3 meetings of 60 minutes.	Regulative directions of planning, monitoring, and evaluating problem solving/learning collaborative processes.	**Episode level**: Utterances were grouped into episodes, based on content, into cognitive episodes, regulation of collaboration episodes, and regulation of learning task episodes.	**Group (positive)**: Cognitive episodes appeared more in the treatment groups compared to control groups. The treatment groups also produced better models. Correlation between cognitive episodes with quality of dynamic model was 0.39 for the experimental group and 0.64 for the control groups. However, regulation of collaboration episodes correlated negatively with quality of dynamic model (in the control group, r=.-66), and regulation of learning task episodes correlated positively (r=.81) in the control groups, but negatively in the experimental groups (r=.-59).
(Chiu, Huang, & Chang, 2000)	Constructing a concept map on "central processing unit" with 11 given concepts. Only one member of the triad can manipulate concept map. Triads can also request a feedback score of their concept map.	12 triads (university students). Single session of 80 minutes.	Collaborative, networked concept mapping tool, plus a scoring facility (comparison of group concept map with expert concept map).	**Utterance level**: Grouped into cooperation (simple and complex), helping (explaining, answering, and informing), and social event behaviors.	**Group (positive)**: The following interactions were correlated with quality of group concept map: number of chat utterance (r=.851), on-task utterances (r=.799), cooperation utterances (r=.754), high-level interaction (r=.872), and complex cooperation (r=.872).