



Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions

Nia M. M. Dowell¹ · Tristan M. Nixon² · Arthur C. Graesser³

Published online: 17 August 2018
© Psychonomic Society, Inc. 2018

Abstract

Roles are one of the most important concepts in understanding human sociocognitive behavior. During group interactions, members take on different roles within the discussion. Roles have distinct patterns of behavioral engagement (i.e., active or passive, leading or following), contribution characteristics (i.e., providing new information or echoing given material), and social orientation (i.e., individual or group). Different combinations of roles can produce characteristically different group outcomes, and thus can be either less or more productive with regard to collective goals. In online collaborative-learning environments, this can lead to better or worse learning outcomes for the individual participants. In this study, we propose and validate a novel approach for detecting emergent roles from participants' contributions and patterns of interaction. Specifically, we developed a group communication analysis (GCA) by combining automated computational linguistic techniques with analyses of the sequential interactions of online group communication. GCA was applied to three large collaborative interaction datasets (participant $N = 2,429$, group $N = 3,598$). Cluster analyses and linear mixed-effects modeling were used to assess the validity of the GCA approach and the influence of learner roles on student and group performance. The results indicated that participants' patterns of linguistic coordination and cohesion are representative of the roles that individuals play in collaborative discussions. More broadly, GCA provides a framework for researchers to explore the micro intra- and interpersonal patterns associated with participants' roles and the sociocognitive processes related to successful collaboration.

Keywords Group communication analysis · Social roles · Group interaction · Computational linguistics

There are many faces to our social identity. When individuals engage in everyday interactions, they act in ways that are both enabled and constrained by social structure: the social context, history, structures of interaction, and attributes that individuals bring to the interaction (Gleave, Welser, Lento, & Smith, 2009; Hare, 1994; Sapru & Boulard, 2015). In this context, social roles provide a valuable window into the underlying sociocognitive structure of group interaction, and one that

researchers can use to differentiate individuals and explain the consequences of an individual's and of an overall group's behavior (Gleave et al., 2009; Mudrack & Farrell, 1995).

The concept of social roles has garnered significant interdisciplinary attention across several areas, including education, social computing, and social and organizational psychology. This has produced a burgeoning literature on social roles in a variety of domains, including teams (Driskell, Driskell, Burke, & Salas, 2017), workplace meetings (Sapru & Boulard, 2015), and collaborative interactions (Strijbos & De Laat, 2010). Roles have been defined more strictly as stated functions and/or responsibilities that guide individual behavior, and as the behavioral patterns exemplified by individuals in social contexts (Chiu, 2000; Hare, 1994; Volet, Vauras, Salo, & Khosa, 2017). This definition is reflected in the two prominent perspectives on roles that appear in the sociological and psychological literatures. The first emphasizes the behaviors associated with a specific appointment in a group or organization. The most obvious examples falling in this class are formal appointments, including employment positions, political offices, military ranks, academic degrees, and

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-018-1102-z>) contains supplementary material, which is available to authorized users.

✉ Nia M. M. Dowell
ndowell@umich.edu

¹ School of Information, University of Michigan, North Quad 4432, 105 S. State Street, Ann Arbor, MI 48109, USA

² Ann Arbor, MI, USA

³ University of Memphis, Memphis, TN, USA

other formal titles. However, this category also includes roles in more ad hoc social situations, such as those that are explicitly teacher-assigned for the purposes of some exercise, or implicitly embodied through prescribed interactions. In this context, the role is a position to which a person is assigned and then performs the behavior associated with that position (Salazar, 1996). The second perspective, by contrast, considers roles as a product of a specific interaction context, consisting of patterns in the sociocognitive behaviors enacted by people (Gleave et al., 2009). These roles are *emergent* in that they develop naturally out of the interpersonal interaction, without any prior instruction or assignment, and are defined (characterized) by their behavioral proximity (similarities and differences) to other interactional partners.

Several studies have emphasized the importance of roles in group interactions (Dillenbourg, 1999; Hoadley, 2010; Jahnke, 2010; Marcos-Garcia, Martinez-Mones, & Dimitriadis, 2015; Sarmiento & Shumar, 2010; Risser & Bottoms, 2014; Spada, 2010; Stahl, Law, Cress, & Ludvigsen, 2014; Strijbos & De Laat, 2010; Volet et al., 2017). Recent work on scripted or assigned roles has shown that the assignment of specific roles facilitates collaborative awareness (Strijbos, Martens, Jochems, & Broers, 2004), team discourse and performance (Gervits, Eberhard, & Scheutz, 2016; Xie, Yu, & Bradshaw, 2014), and the depth of knowledge co-construction by the group (De Wever, Keer, Schellens, & Valcke, 2010; Gu, Shao, Guo, & Lim, 2015). Although assigning roles to group members may produce beneficial outcomes, there are concerns. First, the occurrence of potentially dysfunctional group roles has been generally neglected in the literature, with researchers choosing to focus on those roles that are potentially most productive for the group, as opposed to the roles that actually exist (Lehmann-Willenbrock, Beck, & Kauffeld, 2016). Considering the prevalence of dysfunctional group behavior (e.g., the “bad apple phenomenon”; Felps, Mitchell, & Byington, 2006), deepening our understanding of such negative roles and their influence is crucial. This leads to the second concern, regarding what is captured in role assignment research: Simply because someone is assigned a role does not mean the person will not deviate from said role. Are we, then, exploring roles as intended or as enacted (Hoadley, 2010)? Finally, by attempting to restrict an individual to a single role, one inhibits role and group flexibility, which itself has potential advantages. Doing so also disregards the dynamic and interactive way in which roles are created, negotiated, and evolve among group members during social interaction (Hoadley, 2010; Lehmann-Willenbrock et al., 2016; Salazar, 1996).

Researchers have attempted to detect the emergence of roles during online group interactions (Stahl et al., 2014). The majority of these efforts have relied on predefined content analysis coding schemes and complex taxonomies to determine what roles each individual occupies within the group

(e.g., Arvaja & Hämäläinen, 2008; Volet et al., 2017). For instance, Strijbos and De Laat (2010) provided a valuable conceptual framework of roles in group interactions. Their framework distinguishes eight roles. Four of the roles are reserved for large-group interactions: Pillar, Generator, Hanger-On, and Lurker. However, the remaining four are particularly relevant to small-group interactions: Captain, Over-Rider, Free-Rider, and Ghost. The roles are differentiated along two dimensions that cross orientation (individual, group) and effort (low, high). In the present analysis, the Strijbos and De Laat framework helped guide some of the initial conceptualizations of the processes involved in participant roles. However, as will be shown, we adopted an automated methodological approach that afforded several new dimensions of interaction. Particularly, although extensive knowledge has been gleaned from manual content analyses of emergent social roles, several researchers have pointed out the inherent limitations of this approach: These practices tend to obscure the sequential structure, semantic references within group discussion, and situated methods of interaction through which roles emerge (Çakır, Zemel, & Stahl, 2009; Strijbos, Martens, Prins, & Jochems, 2006; Suthers, Dwyer, Medina, & Vatrapu, 2010). Moreover, manual coding methods are no longer a viable option, with the increasing scale of online group interaction data (Daradoumis, Martínez-Monés, & Xhafa, 2006).

The availability of such data represents a golden opportunity to make advances in understanding social roles and role ecologies (Gleave et al., 2009). However, automatic approaches for detecting emergent social roles are still relatively scarce in the field of collaborative interactions. The attempts that have been made have typically relied on social network analysis (SNA; e.g., Capuano, Mangione, Mazzoni, Miranda, & Orciuoli, 2014; Marcos-Garcia et al., 2015; Stuetzer, Koehler, Carley, & Thiem, 2013). In this context, social roles are characterized in terms of behavioral regularities and network attributes, wherein consistent behaviors resulting in persistent or recurrent interactions between individuals in a social group are potential signals of a meaningful social role (Gleave et al., 2009). One of the advantages of such quantitative methods as SNA is that they alleviate the human time requirement and the attendant subjectivity issues inherent in manual content analyses. However, these strictly structural measures have been criticized for being only surface-level, because they do not capture the deeper-level interpersonal sociocognitive and semantic information found in the discourse interaction (Strijbos & Weinberger, 2010). Automated natural language processing techniques could provide a productive path toward automated role detection, by addressing some of these limitations. Specifically, roles emerge and are sustained through interaction (Hare, 1994; Salazar, 1996), and communication is the basis of any interaction. Indeed, a focus on language and communication has proven quite useful in other explorations of group interaction phenomena (Cade, Dowell, Graesser, Tausczik, & Pennebaker, 2014;

Cai et al., 2017; Dowell, Brooks, Kovanović, Joksimović, & Gašević, 2017; Dowell, Brooks, & Poquet, 2018; Dowell, Cade, Tausczik, Pennebaker, & Graesser, 2014; Dowell & Graesser, 2015; Dowell et al., 2015; Graesser, Dowell, & Clewley, 2017; Ho et al., 2016; Joksimović et al., 2015, 2018). As such, language provides a powerful and measurable behavioral signal that can be used to capture the semantic and sociocognitive interaction patterns that characterize emergent roles, as well as to study their influence on the outcomes of group interactions.

Overview of the present research

The present research has two main objectives. The first is to propose an automated methodology, *group communication analysis* (GCA), for detecting emergent roles in group interactions. The GCA combines computational linguistic techniques with sequential interaction analyses of group communication. The GCA captures theoretically relevant sociocognitive processes that can be used to characterize the *social roles* individuals occupy in group interactions. Tracking the communication dynamics during ongoing group interactions can reveal important patterns about how individual and group processes emerge and unfold over time. The second goal of this research is to explore how the individual-level roles and overall group compositions influence both student and group performance during collaborative interactions. The concepts, methods, and ideas presented in this research are at the intersection of collaborative learning, discourse processes, data mining, and learning analytics. This interdisciplinary research approach will hopefully provide insights and help redefine the nature of roles in group interaction. Specifically, the present research includes analyses of two large, collaborative-learning datasets (Traditional CSCL: learner $N = 854$, group $N = 184$; SMOC: learner $N = 1,713$, group $N = 3,297$) and of one collaborative problem-solving dataset (Land Science: learner $N = 38$, group $N = 630$) to address the research questions outlined below. Although this investigation takes place in the context of collaborative learning, the methodology is flexible and could be applied in any computer-mediated social interaction space that involves linguistic interactions between participants.

Research questions

1. Can individual roles be identified through patterns of communication and participation during collaborative interactions of some specific type or context? We use three approaches to evaluate this research question: (i) comparison to the prior literature, (ii) extensive validation checks, and (iii) assessing the influence of roles on individual and group outcomes.

2a. Do the patterns, if any, observed from Research Question 1 generalize meaningfully to other collaborative interactions of the same type or context?

2b. Do the patterns, if any, observed from Research Question 1 generalize meaningfully to other collaborative interactions of different types or contexts?

3a. How does an individual's role influence individual and group performance?

3b. How does group role diversity and composition influence individual and group performance?

The subsequent sections of the article are organized as follows. First, we provide the theoretical foundation for the GCA measures, followed by a detailed technical description of the construction of the measures. We then move into the methodological features of the present investigation, followed by the details of the cluster analysis that was used to identify specific individual roles in the communication patterns during collaborative interactions. Next, we discuss the linear mixed-effects modeling used to assess the validity of the GCA approach and the influence of roles on individual and group performance. We conclude the article with a detailed discussion of the results in the context of theory, as well as a general discussion of the theoretical, methodological, and practical implications for group interaction research.

Group communication analysis (GCA)

Theoretical motivation for the GCA measures

Social and cognitive processes are the fabric of collaborative learning. The ultimate goal of collaborative learning is the co-constructed knowledge that results from the sharing of information in groups during collaborative tasks (Alavi & Dufner, 2004; Dillenbourg & Fischer, 2007). Learning as a social process is supported by several theoretical perspectives including the social cognitive theory (Bandura, 1994), social-constructivist framework (Doise, 1990), socio-cultural framework (Vygotsky, 1978), group cognition models (Stahl, 2005), situated cognition theory (Lave & Wenger, 1991), and connectivism (Siemens, 2005). Research on the sociocognitive aspects of computer-supported collaborative learning (CSCL) have noted some of the important mechanisms (e.g., social presence, explanation, negotiation, monitoring, grounding, and regulating) and processes (e.g., convergence, knowledge co-construction, meaning-making) that facilitate successful outcomes (Dillenbourg, Järvelä, & Fischer, 2009).

The GCA framework incorporates definitions and theoretical constructs that are based on research and best practices from several areas in which group interaction and collaborative skills have been assessed. These areas include computer-supported cooperative work, team discourse analysis, knowledge sharing, individual problem solving, organizational psychology, and assessment in

work contexts (e.g., military teams, corporate leadership). The framework further incorporates information from existing assessments that can inform the investigation of social roles, including the PISA 2015 CPS Assessment. Despite differences in orientation between the disciplines in which these frameworks have originated, the conversational behaviors that have been identified as valuable are quite similar. The following sections review the theoretical perspectives and sociocognitive processes that were the foundation of the GCA framework and the resulting metrics (i.e., Participation, Internal Cohesion, Responsivity, Social Impact, Newness, and Communication Density). In the presentation of the theoretical principles and sociocognitive processes supporting the GCA metrics, empirical findings are presented whenever possible as illustration and initial support. Table 1 provides a summary of the alignment of the GCA dimensions with their associated theoretical and empirical support.

Participation

Participation is obviously a minimum requirement for collaborative interaction. It signifies a willingness and readiness for participants to externalize and share information and thoughts (Care, Esther, Scoular, & Griffin, 2016; Hesse, Care, Buder, Sassenberg, & Griffin, 2015). Previous research has confirmed that participation, measured as interaction with peers and teachers, has a beneficial influence on

perceived and actual learning, retention rates, learner satisfaction, social capital, and reflection (Hew, Cheung, & Ng, 2010; see Hrastinski, 2008, for a review). Within collaborative groups, individual students who withdraw their participation from group discussion or only participate minimally can undermine learning, either because of lost opportunities for collaboration or by provoking whole-group disengagement (Van den Bossche, Gijsselaers, Segers, & Kirschner, 2006). In CSCL research, typical measures of student participation include the number of a student's contributions (Lipponen, Rahikainen, Lallimo, & Hakkarainen, 2003), the length of posts in online environments (Guzdial & Turns, 2000), or whether contributions are more social (i.e., off-task) rather than focused on content ideas (Stahl, 2000). More recently, Wise, Speer, Marbouti, and Hsiao (2012) argued that a more complete conception of participation in online discussions requires attention not only to participants' overt activity in producing contributions, but also to the less public activity of interacting with the contributions of others, which they have termed "online listening behavior" (Wise et al., 2012). Taken together, this research highlights how individual participants may vary in the amount, type, and quality of participation within a group. Therefore, participation is an important metric to characterize the social roles participants occupy during group interactions. In the present research, *participation* is conceptualized as a necessary, but not a sufficient, sociocognitive metric for characterizing the participants' social roles.

Table 1 Alignment of GCA dimensions with theoretical and empirical support

GCA Dimensions	Psychological & Discursive Processes	Description/ Example Behavioral Makers	Relevant Theoretical Frameworks & Constructs	Empirical Evidence/Theoretical Support
Participation	Engagement	General level of an individual's participation, irrespective of the semantic content of this participation	Activity theory; Social presence; Socio-constructivist	Hesse et al., 2015; Hrastinski, 2008; Hew et al., 2010
Internal cohesion	Monitoring and reflecting	The tendency of an individual to consistency or novelty in their own contributions.	Self-regulation and metacognitive processes	Chan, 2012; Zimmerman, 2001; Barron, 2000; OECD, 2013
Responsivity	Uptake and transactivity	The tendency of an individual to respond, or not, to the contributions of their collaborative peers	Meaning-making; Co-regulation; Interactive alignment; Social coordination; Knowledge building; Common ground; Knowledge convergence	Berkowitz & Gibbs, 1983; Teasley, 1997; Hesse et al., 2015; Suthers, 2006; Volet et al. 2009; Hesse et al., 2015
Social impact	Productive or popular communication	The tendency of a participant to evoke responses, or not, from their collaborative peers	Social coordination; Knowledge building; Common ground; Co-construction	Volet et al. 2009; Hesse et al., 2015; Suthers, 2006
Newness	Novelty of information shared	The tendency to provide new information or to echo previously stated information, irrespective of the originator of the information.	Monitoring; Information sharing	Chi, 2009; Hesse et al., 2014; Mesmer-Magnus & Dechurch, 2009; Kirschner et al., 2008
Communication density	Concise communication	The extent to which participants convey information in a concise manner	Common ground; Effective communication	Gorman et al., 2003, 2004

Internal cohesion, responsivity, and social impact

Simply placing participants in groups does not guarantee collaboration or learning (Kreijns, Kirschner, & Jochems, 2003). For collaboration to be effective, participants must participate in shared knowledge construction, have the ability to coordinate different perspectives, commit to joint goals, and evaluate their collective activities together (Akkerman et al., 2007; Beers, Boshuizen, Kirschner, & Gijsselaers, 2007; Blumenfeld, Kempler, & Krajcik, 2006; Fiore & Schooler, 2004; Kirschner, Paas, & Kirschner, 2009; Roschelle & Teasley, 1995). This raises an important question that has been reoccurring theme in the CSCL literature: What makes collaborative discourse productive for learning? (Stahl & Rosé, 2013). This question has been studied with a related focus and comparable results across several CSCL subcommunities.

Collaborative knowledge construction is understood as an unequivocally interpersonal and contextual phenomenon, but the role of an individual interacting with themselves should also be taken into account (Stahl, 2002). Successful collaboration requires that each individual *monitor* and *reflect* on their own knowledge and contributions to the group (Barron, 2000; OECD, 2013). This points to the importance of self-regulation in collaborative interactions (Chan, 2012; Zimmerman, 2001). Self-regulation is described as an active, constructive process in which participants set goals, and monitor and evaluate their cognition, affects, and behavior (Azevedo, Winters, & Moos, 2004; Pintrich, 2000; Winne, 2013). During collaborative interactions, this is necessary for individuals to appropriately build on and integrate their own views with those of the group (Kreijns et al., 2003; OECD, 2013). The process of individuals engaging in self-monitoring and reflection may be reflected in their *internal cohesion*. That is, a participant's current and previous contributions should be, to some extent, semantically related to each other, which can indicate the integration and evolution of thoughts through monitoring and reflecting (i.e., self-regulation). However, overly high levels of internal cohesion might also suggest that a participant is not building on or evolving his or her thoughts, but rather is simply reiterating the same static view. Conversely, very low levels of internal cohesion might indicate that a participant has no consistent perspective on offer to the group, is simply echoing the views of others, or is only engaging at a nominal or surface level. Therefore, we should expect productive roles to exhibit a moderate degree of internal cohesion.

Participants must also monitor and build on the perspectives of their collaborative partners in order to achieve and maintain a shared understanding of the task and its solutions (Dillenbourg & Traum, 2006; Graesser et al., 2016; Hmelo-Silver & Barrows, 2008; OECD, 2013; Stahl & Rosé, 2013). In the CSCL literature this shared understanding has been referred to as knowledge convergence, or common ground (Clark, 1996; Clark & Brennan, 1991; Fiore & Schooler,

2004; Roschelle, 1992). It is achieved through communication and interaction, such as building a shared representation of the meaning of the goal, coordinating efforts, understanding the abilities and viewpoints of group members, and mutual monitoring of progress toward the solution. These activities are supported in several collaborative-learning perspectives (e.g., cognitive elaboration, Chi, 2009; sociocognitive conflict, Doise, 1990; Piaget, 1993; co-construction, Hatano, 1993; Van Boxtel, 2004), each of which stresses different mechanisms to facilitate learning during group interactions (giving, receiving, and using explanations, resolving conflicts, co-construction). However, all these perspectives are in alignment on the idea that it is the participants' elaborations on one another's contributions that support learning.

These social processes of awareness, monitoring, and regulatory processes all fall under the shared umbrella of co-regulation. Volet, Summers, and Thurman (2009) proposed *co-regulation* as an extension of self-regulation to the group or collaborative context, wherein co-regulation is described as individuals working together as multiple self-regulating agents, all socially monitoring and regulating each other's learning. In a classroom study of collaborative learning using hypermedia, Azevedo et al. (2004) demonstrated that collaborative outcomes were related to the use of regulatory behaviors. In this process, the action of one student does not become a part of the group's common activity until other collaborative partners react to it. If other group members do not react to a student's contribution, this suggests that the contribution was not seen as valuable by the other group members and would be an "ignored co-regulation attempt" (Molenaar, Chiu, Slegers, & van Boxtel, 2011). Therefore, the concepts of *transactivity* and *uptake* (Table 1) in the CSCL literature are important in this context of co-regulation and active learning, in the sense that a student takes up another student's contribution and continues it (Berkowitz & Gibbs, 1983; Suthers, 2006; Teasley, 1997). Students can engage in higher or lower degrees of co-regulation through monitoring and coordinating. These processes will be represented in their discourse.

Monitoring and regulatory processes are, hopefully, externalized during communication with other group members. We can capture the degree to which an individual is monitoring and incorporating the information provided by their peers by examining the semantic relatedness between the individual's current contribution and the previous contributions of their collaborative partners. This measure is called *responsivity* in the present research. For example, if an individual's contributions are, on average, only minimally related to those of peers, we would say this individual has low responsivity. Similarly, we can capture the extent to which a participant's contributions are seen as meaningful or worthy of further discussion (i.e., *uptake*) among peers by measuring the semantic relatedness between the participant's current contribution and those that follow from their collaborative partners. This measure is

called *social impact* in the present research. Participants have high social impact to the extent that their contributions are often semantically related to the subsequent contributions from the other collaborative group members.

The collaborative-learning literature highlights the value of students clearly articulating arguments and ideas, as well as elaborating on and making connections between contributions. For instance, Rosé and colleagues' work has concentrated explicitly on such properties as transactivity (Gweon, Jain, McDonough, Raj, & Rosé, 2013; Joshi & Rosé, 2007; Rosé et al., 2008), as well as the social aspects and conversational characteristics that facilitate the recognition of transactivity (Howley & Mayfield, 2011; Howley, Mayfield, & Rosé, 2013a; Howley, Mayfield, Rosé, & Strijbos, 2013b; Wen, Yang, & Rosé, 2014). Their research adopts a sociocognitive view (Howley, Mayfield, Rosé, & Strijbos, 2013) that emphasizes the significance of publically articulating ideas and encouraging participants to listen carefully to and build on one another's ideas. Participants engaging in this type of activity have the chance to notice discrepancies between their own mental model and those of other members of the group. The discussion provides opportunities to engage in productive cognitive conflict and knowledge construction (Howley, Mayfield, Rosé, & Strijbos, 2013a). Additionally, participants benefit socially and personally from the opportunity to take ownership over ideas and position themselves as valuable sources of knowledge within the collaborative group (Howley & Mayfield, 2011).

Newness and communication density

For collaboration to be successful, participants must engage in effective information sharing. Indeed, one of the primary advantages of collaborative interactions and teams is that they provide the opportunity to expand the pool of available information, thereby enabling groups to reach higher quality solutions than could be reached by any one individual (Hesse et al., 2015; Kirschner, Beers, Boshuizen, & Gijsselaers, 2008; Mesmer-Magnus & Dechurch, 2009). However, despite the intuitive importance of effective information sharing, a consistent finding from research is that groups predominantly discuss information that is *shared* (and so known to all participants) at the expense of information that is *unshared* (known to a single member) (Dennis, 1996; Stasser & Titus, 1985; see Wittenbaum & Stasser, 1996, for a review). This finding has been called *bias information sharing* or *bias information pooling* in the collective information-sharing paradigm. It shares some similarities with the groupthink phenomena (Janis, 1983), which is the tendency for groups to drive for consensus that overrides critical appraisal and the consideration of alternatives. The collective preference for redundant information can detrimentally affect the quality of the group interactions (Hesse et al., 2015) and decisions made within the

group (Wittenbaum, Hollingshead, & Botero, 2004). However, collaborative interactions benefit when the members engage in the constructive discourse of inferring and sharing new information and integrating new information with existing prior knowledge during the interaction (Chi, 2009; Chi & Menekse, 2015).

The distinction between given (old) information versus new information in discourse is a foundational distinction in theories of discourse processing (Haviland & Clark, 1974; Prince, 1981). *Given* information includes words, concepts, and ideas that have already been mentioned in the discourse; *new* information involves words, concepts, and ideas that have not yet been mentioned, which builds on the given information or launches a new thread of ideas. In the present research, the extent to which learners provide new information rather than referring to previously shared information will be captured by a measure called *newness*.

In addition to information sharing, the team performance literature also advocates for concise communication between group members (Gorman, Cooke, & Kiekel, 2004; Gorman, Foltz, Kiekel, Martin, & Cooke, 2003). This is one of the reasons that formal teams, like military units, typically adopt conventionalized terminology and standardized patterns of communication (Salas, Rosen, Burke, Nicholson, & Howse, 2007). It is suggested that this concise communication is possible when there is more common ground within the team and when shared mental models of the task and team interaction are present (Klein, Feltovich, Bradshaw, & Woods, 2005). The *communication density* measure used in the present research was first introduced by Gorman et al. (2003) in team communication analysis, to measure the extent to which a team conveys information in a concise manner. Specifically, the rate of meaningful discourse is defined by the ratio of semantic content to number of words used to convey that content. Using this measure, we will be able to further characterize the social roles that participants take on during collaborative interactions.

Taken together, we see that the sociocognitive processes involved in collaboration are internal to the individual but they are also manifested in the interactions with others in the group (Stahl, 2010). In particular, during group interactions, participants need to self-regulate their own learning and contributions, and co-regulate the learning and contributions of their collaborative partners. Reciprocally, the discourse of group members influences each participants' own monitoring and cognition (Chan, 2012; Järvelä, Hurme, & Järvelä, 2011). The social roles explored in this research are not necessarily reducible to processes of individual minds nor do they imply the existence of some sort of group mind. Rather, they are characterized by and emerge from the sequential interaction and weaving of semantic relations within a group discourse. The artifacts of transcribed communication resulting from collaborative interactions provides a window into the cognitive and social processes that define the participants' social roles. Thus, communication among the group

members can be assessed to provide measures of participation, social impact, internal cohesion, responsiveness, newness, and communication density. These measures, which make up the GCA framework, define a space that can encompass many key attributes of a collaborative group interaction. Because participants exhibit more or less internal cohesion, responsiveness, social impact, new information, and communication density, we can associate them with a unique point in that space. As we will show, these points tend to cluster into distinct regions, corresponding to distinct patterns in behavioral engagement style and contribution characteristics. As such, these clusters represent characteristic roles that individuals take on during collaborative interactions, and they have a substantial impact on the overall success of those interactions.

Construction of the GCA and group performance measures

Transcripts of online group conversations provide two principal types of data (Foltz & Martin, 2009). First, transcript metadata describe the patterns of interactions among group members. This includes who the participants are, as well as the number, timing, and volume of each of their contributions. Second is the actual textual content of each individual contribution, from which we can calculate the semantic relationships among the contributions. This involves taking semi-unstructured log file data, as depicted in Fig. 1, and performing various transformations in order to infer the semantic relationships among the individual contributions, as depicted in Fig. 2.

Conversations, including collaborative discussions, commonly follow a statement–response structure, in which new statements are made in response to previous statements (i.e., uptake; Suthers, 2010) and subsequently trigger further statements in response (see Fig. 2). The structures of different online communications and discussion systems provide different affordances to the analyst to attribute a specific contribution as a response to some prior contribution. Regardless of the structure of the system, participants may, in a single contribution, refer to concepts and content presented in multiple previous contributions, made throughout the conversation, by either themselves or other group members. Thus, a single contribution may be in response, to varying degrees, to many previous contributions, and it may in turn trigger, to varying degrees, multiple subsequent responses.

The analytical approach of the GCA was inspired by analogy to the cross- and auto-correlation measures from time-series analysis. Standard correlation measures how likely two variables are to be related. Cross-correlation similarly measures the relatedness between two variables, but with a given interval of time (or lag) between them. That is, for variables x and y , and a lag of τ , the cross-correlation would be the correlation of $x(t)$ with $y(t + \tau)$, across all applicable times, t , in the time series. Standard correlation can be seen as a special case of cross-correlation for which $\tau = 0$, and auto-correlation as a special case of one variable being correlated with itself, shifted in time ($\tau \neq 0$). By plotting the values of a cross-correlation at different values of τ (typically from 1 up to some reasonably large value), one can identify whether there are any statistically significant time-dependent relationships between the variables being examined. Such cross-correlation plots are commonly used in the qualitative exploration of time-series data.

group_id	person_id	chat_time	Contribution_number	chat_text
958	A	5:12:00 PM	0	hello
958	B	5:13:00 PM	1	Hi! Ok cool, everyone's here. So first question
958	D	5:13:00 PM	2	Hello! So the certain characteristics to be considered to have a personality disorder?
958	C	5:14:00 PM	3	okay so, for the first question, certain characteristics: doesn't it have to be like a stable thing?
958	E	5:14:00 PM	4	i think the main thing about having a disorder is that its disruptive socially and/or makes the person a danger to himself or others
958	A	5:14:00 PM	5	yes, it is more fixed
958	A	5:14:00 PM	6	or, stable over time
958	B	5:14:00 PM	7	Great, thanks, and it also mentioned it can't be because of drugs
958	C	5:15:00 PM	8	They have to have like unrealistic fantasies
958	B	5:15:00 PM	9	yeah and not normal in their culture
958	B	5:15:00 PM	10	no drugs or physical injury
958	D	5:15:00 PM	11	i think that covers them? haha
958	C	5:15:00 PM	12	Right, it can't be from drugs and it begins in early adulthood or adolescence
958	B	5:15:00 PM	13	ok, so arrogance doesn't just define it, they have to have most of these characteristics
958	A	5:15:00 PM	14	Yes, early adulthood, ok i think we got them
958	C	5:15:00 PM	15	is it most or is it like 6?
958	E	5:16:00 PM	16	Think that covered them
958	B	5:16:00 PM	17	they mention six, and we have got them
958	B	5:16:00 PM	18	Okay so second question?

Fig. 1 Depiction of semi-unstructured log file data, a typical artifact of CSCL interactions.

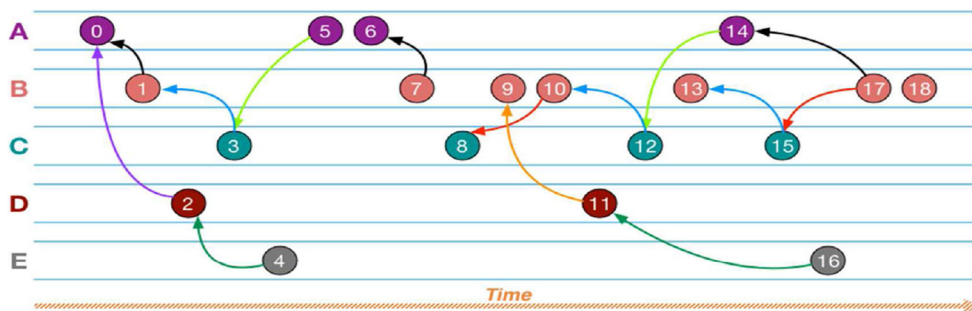


Fig. 2 Schematic representation of inferring the semantic relationships among students’ contributions in group interactions. The letters (i.e., A, B, C, D, E) on the vertical axis refer to students within a group interaction, and the numbers represent the sequential order of their discourse contributions.

Although we might apply standard auto- and cross-correlation in order to examine the temporal patterns in *when* participants contribute, we are primarily interested in understanding the temporal dynamics of *what* they contribute, and in what the evolution of the conversation’s semantics can teach us about the group’s collaboration. To this end, a fine-grained measure of the similarity of participants’ contributions is needed to capture the multiresponsive and social impact dynamics that may be present in collaborative interactions. There are different techniques for calculating the semantic similarity between two contributions. Two popular methods are content word overlap (CWO) and latent semantic analysis (LSA). Each has its own strengths and weaknesses (Hu, Cai, Wiemer-Hastings, Graesser, & McNamara, 2007); however, these methods typically produce comparable results. In this research, similarity is measured using LSA. The semantic cohesion of contributions at fixed lags in the conversations can be computed much in the same way that cross-correlation evaluates correlation between lagged variables. Various aggregations of this auto- and cross-cohesion form the basis of the GCA’s responsivity measures.

In addition to the GCA measures, the identification of topics covered in the group discussion affords us an objective measure of the overall group performance that is independent of the individual student performance (i.e., pre- and post-test scores). In the sections below, we describe the technical details of the construction of each of the GCA measures and of the group performance measure (i.e., topic relevance).

Participation measures

The chat logs of a group discussion can be thought of as a sequence of individual contributions (i.e., verbal expressions within a conversational turn). In this sense, the boundaries of a contribution are defined by the nature of the technology that mediates the group discussion. A single contribution is a single message transmitted from one user to a group of others by way of a messaging service,

or a single posting by a single user to a discussion forum. There may be multiple speech acts within a single contribution, but these will be treated as a single contribution. Furthermore, a single user may transmit further contributions, immediately subsequent to their first, but these will be treated as separate contributions. So, the primary unit of analysis is a single contribution from a single user.

Let C represent the sequence of contributions, with c_t representing the t^{th} contribution in the sequence. Let $n = |C|$ denote the length of the sequence. Since contributions represent turns in the discussion over time, the variable t will be used to index individual contributions and will also be referred to as “time.” The values of t will range from 1 to n :

$$t \in \mathbb{Z}; 1 \leq t \leq n \tag{1}$$

Let P be the set of participants in the discussion, of size $k = |P|$. Variables a and b in the following will be used to refer to arbitrary members (participants) in this set. To identify the contributor (or participant) who originated each statement, we define the following participation function, as depicted in Eq. 2:

$$p_a(t) = \begin{cases} 1, & \text{if contribution } c_t \text{ was made by participant } a \in P \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The participation function for any participant, a , effectively defines a sequence,

$$P_a = \{p_a(t)\}_{t=1}^n = \{p_a(1), p_a(2), p_a(3), \dots, p_a(n)\} \tag{3}$$

of the same length, n , as the sequence of contributions C , which has the value 1 whenever participant a originated the corresponding contribution in C , and 0 everywhere else. Using this participation function, it is relatively simple to define several useful descriptive measures of participation in the discussion. The number of contributions made by any participant is

$$\|P_a\| = \sum_{t=1}^n p_a(t) \tag{4}$$

The sample mean participation of any participant is the relative proportion of his or her contributions out of the total,

$$\bar{p}_a = \frac{1}{n} \|P_a\| \quad (5)$$

and the sample variance in that participation is

$$\sigma_a^2 = \frac{1}{n-1} \sum_{t=1}^n (p_a(t) - \bar{p}_a)^2 \quad (6)$$

If every participant contributed equally often, say by taking turns in round-robin fashion, then, for every participant,

$$\|P_a\| = \frac{n}{k} \quad (7)$$

This, in turn would result in mean participation scores of

$$\bar{p}_a = \frac{1}{n} \cdot \frac{n}{k} = \frac{1}{k} \quad (8)$$

which will naturally get smaller for larger groups (larger k). To adjust for different levels of equal participation, we use the following measure¹ to characterize individual participation:

$$\hat{p}_a = \bar{p}_a - \frac{1}{k} \quad (9)$$

This gives the mean participation above or below what we might expect from perfectly equal participation. In the case in which every participant contributes equally (Eq. 7), this measure would be 0. A participant who contributed more than the equal-participation amount would have a positive score, one who contributed less, a negative score. We refer to this as the *group-relative mean participation*.

We can, equivalently, represent the sequences of all participant as a $k \times n$ matrix, M , by stacking the k participation sequences as rows, in any arbitrary ordering (such that i is an index over participants). Under this representation, the (i, j) th entry of the matrix is the j th contribution of participant i :

$$M_{ij} = p_i(j); i \in P, 1 \leq j \leq n \quad (10)$$

By the definition of contributions given above, each contribution c_t was originated by one and only one participant, so the participation function, p_t , will take on a value of 1 for exactly 1 participant at each time t , and be 0 for all other

¹ During the review, it was observed that this measure does not correctly put the participation on a common scale across conversations of different sizes. A more correct adjustment would be

$$\hat{p}_a = \frac{\bar{p}_a - \frac{1}{k}}{\frac{1}{k}} = k\bar{p}_a - 1$$

This gives a measure of the individual level of participation, above or below equal participation, in units of equal participation, and is comparable across groups of different sizes.

participants. It follows, therefore, that the sum of each column in the matrix (10) would be exactly 1.

Since each participation sequence is, in effect, a time series of participant contributions, our goal of characterizing the interactions between participants is a problem of characterizing their corresponding participation time series. The field of time-series analysis gives us tools that we can either use directly or adapt to our needs. Specifically, we can make use of the cross-correlation between any two participants a and b :

$$\rho_{a,b}(\tau) = \frac{1}{(n-1)\sigma_a \cdot \sigma_b} \sum_{t=\tau+1}^n p_a(t) \cdot p_b(t-\tau) - n \cdot \bar{p}_a \cdot \bar{p}_b \quad (11)$$

where the variable τ ,

$$\tau \in \mathbb{Z}; \tau \geq 0 \quad (12)$$

is some interval of time (or “lag”) between the initial contribution of b and then some subsequent contribution of a . A lag-1 cross-correlation between two participants will give a measure of how frequently one participant contributes immediately after the other participant. A lag-2 cross-correlation will give a measure of the responsiveness of the one participant after a single intervening contribution. One can qualitatively examine temporal patterns in any pair of participants’ contributions by plotting this function for some reasonable number of lags. By looking at these plots for all pairs of users, one can examine the patterns for the entire group.

Latent semantic analysis

LSA represents the semantic and conceptual meanings of individual words, utterances, texts, and larger stretches of discourse based on the statistical regularities between words in a large corpus of natural language (Landauer, McNamara, Dennis, & Kintsch, 2007). The first step in LSA is to create a word-by-document co-occurrence matrix, in which each row represents a unique word and each column represents a “document” (in practice, this typically means a sentence, paragraph, or section of an actual document). The values of the matrix represent counts of how many occurrences there are of each word in each document. For example, if the word “dog” appears once each in Documents 1 and 9 and twice in Document 50, and is considered the first word in the dataset, then the value of 1 will be in cells (1, 1) and (1, 9), and the value of 2 in cell (1, 50). The occurrence matrix will then be weighted. Each row is weighted by a value indicating the importance of the corresponding word. Functional words (or “stop words”) that occur with nearly even frequency across all documents receive small weights, since they are less useful at distinguishing documents. By contrast, words that have very different occurrences across the documents, and hence indicate more meaningful content terms, get higher weights. The most widely used weighting methods are term-frequency

inverse document-frequency (TF-IDF) and log-entropy. A principal component analysis (PCA) is then performed on the weighted matrix by means of singular-value decomposition (SVD) matrix factorization. PCA is a procedure that allows one to reduce the dimensionality of a set of data such that it minimizes distortions in the relationships of the data. In the context of LSA, PCA allows us to reduce the word-by-document matrix to approximately 100–500 functional dimensions, which represent in compact form the most meaningful semantic relationships between words. The SVD procedure also yields a matrix that can be used to map the words from the original text corpus into vectors in a semantic space described by these semantic dimensions (i.e., LSA space).

When building an appropriate LSA space, it is necessary to have a corpus that broadly covers the topics under investigation. The Touchstone Applied Science Associates (TASA) corpus is a good example of a comprehensive set of tens of thousands of texts across numerous subject areas and spanning a range of levels of complexity (grade levels), which is suitable for building a general semantic space. In some instances, however, researchers desire a more custom corpus covering a specific domain, which was the case in the present research. The source corpora used in this research were conversational transcripts of collaborative interactions that are not large enough to construct an LSA space. Furthermore, these transcripts refer to ideas and concepts that are not explicitly described in the transcripts. To obtain an appropriate representation of the semantic space, we needed to include external material that would cover the topics of the conversations. One way to handle this problem was to enrich the source corpus with additional material that could provide appropriate background knowledge for key terms represented in the conversational transcripts (Cai, Li, Hu, & Graesser, 2016; Hu, Zhang, Lu, Park, & Zhou, 2009). The process began with collecting a “seed” corpus of representative material (Cai, Burkett, Morgan, & Shaffer, 2011). In the present research, this included the chat transcripts for each dataset and the associated assigned reading material for the students. This was done separately for each of the three datasets (described in the Method section), to produce a custom, domain-specific seed corpus. This seed corpus was then scanned for key terms, which were used to scan the Internet for documents (i.e., Wikipedia articles) on the topics mentioned in the seed corpus. The identified documents were used to create an expanded LSA space that was more comprehensive than the underlying transcripts on their own. For details on the extended LSA spaces for each of the corpora used in this research, please see the supplementary material.

By translating text from the corpus into numerical vectors, a researcher can then perform any number of mathematical operations to analyze and quantify the characteristics of the text. One key operation is to compute the semantic similarity between any two segments of text. In the context of interactive chat, the similarity contributions c_t and c_u (where u , like t , is an

index over time), can be computed by first projecting them into the LSA space, yielding corresponding document vectors \vec{d}_t and \vec{d}_u . The projection is done by matching each word or term that occurs in the contribution, and locating the normalized term-vector for that word (calculated by the SVD process). These vectors are added together to get a vector corresponding to the entire contribution. If any term does not occur in the LSA space, it is ignored, and so does not contribute to the resulting vector. However, the construction of the space is such that this is very rare. Then, the cosine similarity of textual coherence (Dong, 2005), is computed on the document vectors \vec{d}_t and \vec{d}_u , as described in equation 13. The cosine similarity ranges from -1 to 1 , with identical contributions having a similarity score of 1 , and completely nonoverlapping contributions (no shared meaning) having a score of 0 or below (although, in practice, negative text similarity cosines are very rare).

$$\cos(\vec{d}_t, \vec{d}_u) = \frac{\vec{d}_t \cdot \vec{d}_u}{\|\vec{d}_t\| \cdot \|\vec{d}_u\|} \quad (13)$$

The primary assumption of LSA is that there is some underlying or “latent” structure in the pattern of word usage across contexts (e.g., turns, paragraphs or sentences within texts), and that the SVD of the word-by-document frequencies will approximate this latent structure. The method produces a high-dimensional semantic space into which we can project participant contributions and measure the semantic similarity between them.

Using this LSA representation, students’ contributions during collaborative interactions may be compared against each other in order to determine their semantic relatedness, and additionally, assessed for magnitude or salience within the high-dimensional space (Gorman et al., 2003). When used to model discourse cohesion, LSA tracks the overlap and transitions of meaning of text segments throughout the discourse.

Using this semantic relatedness approach, the semantic similarity score of any pair of contributions can be calculated as the cosine of the LSA document vectors corresponding to each contribution. This works well as a measure of similarity between pairs of contributions. However, it must be aligned with the participation function in order to get a measure of the relationship between those participants in the discussion. As we demonstrated above, the participation function can be used to select pairs of contributions related to a specific participant-participant interaction, and will screen out all other pairs of interactions. We therefore define a semantic similarity function:

$$s_{ab}(t, u) = p_a(t) \cdot p_b(u) \cdot \cos(\vec{d}_t, \vec{d}_u) \quad (14)$$

This represents the semantic similarity for contributions c_t and c_u only when contribution c_t was made by participant a and

c_u was made by participant b ; otherwise, it is 0 [because, in this case, either $p_a(t)$ or $p_b(u)$, or both, would be 0]. This product will form the foundation of several novel measures to characterize different aspects of participant involvement in the group discussion: the general participation, responsiveness, internal cohesion, and social impact. These measures, described below, will be aligned and compared with Strijbos and De Laat's (2010) conceptual framework to identify participants' roles.

Cross-cohesion

This measure is similar in construction to the cross-correlation function, in that it assesses the relatedness of two temporal series of data to each other, at a given lag τ , though it relies on semantic cohesion rather than correlation as the fundamental measure of relatedness. This measure captures how responsive one participant's contributions are to another's over the course the collaborative interactions. Cross-cohesion is defined by averaging the semantic similarity of the contributions of the one participant to those of the other when they are lagged by some fixed amount, τ , across all contributions:

$$\xi_{ab}(\tau) = \begin{cases} 0, & \|P_{ab}(\tau)\| = 0 \\ \frac{1}{\|P_{ab}(\tau)\|} \sum_{t=\tau+1}^n s_{ab}(t-\tau, t), & \|P_{ab}(\tau)\| \neq 0 \end{cases} \quad (15)$$

It is normalized by the total number of τ -lagged contributions between the two participants, as expressed in Eq. 16:

$$\|P_{ab}(\tau)\| = \sum_{t=\tau+1}^n p_a(t-\tau) \cdot p_b(t) \quad (16)$$

We use the Greek letter ξ (xi) to signify the cross-cohesion function. We refer to $\xi_{ab}(\tau)$ as the “cross-cohesion of a to b at τ ” or as the “ τ -lagged cross-cohesion of a to b .” The cross-cohesion function measures the average semantic similarity of all τ -lagged contributions between two participants. As such, it gives an insight into both the degree and rapidity to which one participant may be responding to the comments of another. The first participant, denoted by a , is that user whose prior contribution potentially has influenced a subsequent comment. The second participant, denoted by b , is that user whose contribution potentially responds to some part of the initiator's contribution. In this way, cross-cohesion can give a measure of the average semantic uptake between participants at a given time-scale. The cross-cohesion at 1 represents the degree of uptake observed in the immediately previous contribution. The propensity for uptake to contributions after one intervening contribution is characterized by the 2-lagged cross-cohesion matrix, and so on. In the special case that the first and the second participants are the same, we may refer to this as the *autocohesion function*, in similar fashion to the autocorrelation function. The autocohesion function measures consistency over time in the semantics of a single participant's contributions. The most similar work to date (Samsonovich,

2014) has made use of the standard cross-correlation function applied to time series of numeric measures computed from natural language, and then draws inferences from these as to the nature of social interactions. However, the use of semantic similarity as the base measure of relatedness, in lieu of correlation, is, to our knowledge, entirely novel.

Responsivity

Cross-cohesion at a single lag may not be very insightful on its own, in that it represents a very narrow slice of interaction. By averaging over a wider window of contributions, we can get a broader sense of the interaction dynamics between the participants.

For a conversation with $k = |P|$ participants, and given some arbitrary ordering of participants in P , we can represent cross-cohesion as a $k \times k$ matrix $X(\tau)$, such that the element in row i , column j is given by the cross-cohesion function $\xi_{ij}(\tau)$. We refer to this matrix as “ τ -lagged cross-cohesion matrix,” or “cross-cohesion at τ .” The rows of the matrix represent the responding students, who we refer to as the respondents. The columns of the matrix represent the initiating participants, referred to as the initiators. We define responsivity across a time window as follows:

$$R(w) = \frac{1}{w} \sum_{\tau=1}^w X(\tau) \quad (17)$$

This will be referred to this as “ w -spanning responsivity” or “responsivity across w .” An individual entry in the matrix, $r_{ab}(w)$ is the “ w -spanning responsivity of a to b ” or the “responsivity of a to b across w .” These measures form a moving average of responsivity across the entire dialogue. The window for the average consists of a trailing subset of contributions, starting with the most current and looking backward over a maximum of w prior contributions. The characteristics of an individual participant can be obtained by averaging over their corresponding rows or columns of the w -spanning responsivity matrix, and by taking their corresponding entry in the diagonal of the matrix. For details on the spanning window calibration used for the datasets in the present research, please see the supplementary material.

Internal cohesion Internal cohesion is the measure of how semantically similar a participant's contributions are with their own previous contributions during the interaction. The participant's “ w -spanning internal cohesion” is characterized by the corresponding diagonal entry in the w -spanning responsivity matrix:

$$r_{aa}(w) \quad (18)$$

The internal cohesion is effectively the average of the autocohesion function of the specified participant over the first w lags.

Overall responsivity Each row in the w -spanning responsivity matrix is a vector representing how the corresponding participant has responded to all others. To characterize how responsive a participant is to all other group members' contributions during the collaborative interactions, we take the mean of these row vectors (excluding the participant of interest):

$$\bar{r}_a(w) = \frac{1}{k-1} \sum_{i=1; i \neq a}^k r_{ai}(w) \quad (19)$$

This is referred to as the “ w -spanning responsivity of a ,” or just the “overall responsivity of a ,” for short.

Social impact Each column in the w -spanning responsivity matrix is a vector representing how contributions initiated by the corresponding participant have triggered follow-up responses. In a similar fashion to the overall responsivity described above, a measure of each individual participant's social impact can be calculated by averaging over these column-vectors (excluding the participant of interest):

$$\bar{i}_a(w) = \frac{1}{k-1} \sum_{j=1; j \neq a}^k r_{ja}(w) \quad (20)$$

This is referred to as the “ w -spanning impact of a ,” or just the “social impact of a ,” for short.

LSA given–new Participants' contributions can vary in how much *new* versus *given* information they contain (Hempelman et al., 2005; McCarthy et al., 2012). Note that, for the purposes of the present research, we were more interested in a measure of the amount of new rather than given information provided by participants. This is motivated by the fact the responsivity measures already capture the social equivalent of “givenness,” which is more relevant in the contexts of group interactions. Establishing how much new information is provided in any given contribution can be meaningful to the dynamics of the conversation, as well as to characterize the ways in which different participants contribute. Following the method of Hu et al. (2003), the given information at the time of contribution t is a subspace of the LSA spanned by the document vectors of all previous contributions:

$$G_t = \text{span}\{\vec{d}_1, \vec{d}_2, \dots, \vec{d}_{t-1}\} \quad (21)$$

The semantic content of the current contribution can then be divided into the portion already given by projecting the LSA document vector for the current contribution onto the subspace defined in Eq. 22:

$$\vec{g}_t = \text{Proj}_{G_t}(\vec{d}_t) \quad (22)$$

There is also the portion of semantic content that is new to the discourse, which we can explore by projecting the same

document vector onto the orthogonal complement of the given subspace, as defined in Eq. 23:

$$\vec{n}_t = \text{Proj}_{G_t^\perp}(\vec{d}_t) \quad (23)$$

This is the portion perpendicular to the given subspace. Of course, the semantic content of the contribution is completely partitioned by these projections, so

$$\vec{d}_t = \vec{g}_t + \vec{n}_t \quad (24)$$

To get a useful measure of the total amount of new semantic content provided in any given contribution, we take the relative proportion of the size of the new vector to the total content provided:

$$n(c_t) = \frac{\|\vec{n}_t\|}{\|\vec{n}_t\| + \|\vec{g}_t\|} \quad (25)$$

This given–new value ranges between 0 (*all given content, nothing new*) to 1 (*all new content*).

Newness We can characterize the relative new content provided by each individual participant by averaging over the given–new scores of this participant's contributions:

$$\bar{N}_a = \frac{1}{\|P_a\|} \sum_{t=1}^n p_a(t) \cdot n(c_t) \quad (26)$$

Communication density Another meaningful measure involves calculating the average amount of semantically meaningful information provided in a contribution. This measure was first established by Gorman et al. (2003) in their work examining team communication in a synthetic military aviation task. This measure differs from the *given–new* measure in that it is entirely calculated from the contribution c_i and its corresponding LSA vector, \vec{d}_i , and does not consider any prior contributions. The communication density is defined in Eq. 27:

$$D_i = \frac{\|d_i\|}{\|c_i\|} \quad (27)$$

where $\|d_i\|$ is the norm of the LSA vector and $\|c_i\|$ is the length of the contribution in words. Thus, communication density gives the per-word amount of semantic meaning for any contribution. To characterize the communication density of a particular participant, we must calculate the average density over all this participant's contributions:

$$\bar{D}_a = \frac{1}{\|P_a\|} \sum_{t=1}^n p_a(t) \cdot D_t \quad (28)$$

The six measures that comprise the GCA are summarized in Table 2.

Table 2 Collaborative interaction process measures from the GCA

Measure	Equation	Description
Participation	(9)	Mean participation of any participant above or below what you would expect from equal participation in a group of the size of theirs
Overall responsiveness	(19)	Measure of how responsive a participant's contributions are to all other group members' recent contributions
Internal cohesion	(18)	How semantically similar a participant's contributions are with the participant's own recent contributions
Social impact	(20)	Measure of how contributions initiated by the corresponding participant have triggered follow-up responses
Newness	(26)	The amount of new information a participant provides, on average
Communication density	(28)	The amount of semantically meaningful information

Topic modeling

The cohesion-based discourse measures described above capture important intrapersonal and interpersonal dynamics, but an additional data-mining technique is needed to capture the themes and topics of the collaborative discussions. The identification of covered topics is of particular interest for the present analyses, because it affords an assessment of the overall group performance that is independent of the individual student performance (i.e., pretest and posttest scores). Latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003), more commonly known as *topic modeling* (Steyvers & Griffiths, 2007), is a method of deriving an underlying set of topics from an unlabeled corpus of text.

Topic modeling allows researchers to discover the common themes in a large body of text and to what extent those themes are present in individual documents. Topic modeling has frequently been used to explore collaborative-learning contexts (e.g., Cai et al., 2017). In this research, LDA topic models were used to provide an inference mechanism of underlying topic structures through a generative probabilistic process. This generative process delivers a distribution over topics for each document in the form of a proportion. This distribution can be used to find the topics most representative of the contents of that document. These distributions can also be considered as data for future analyses, as every document's distribution describes a document-topic "fingerprint." For this research, the topic model corpus for each of the three datasets (described in the Method section) consisted of the extended corpora produced with the "seed method" described earlier (see the Latent Semantic Analysis section above). A topic model was then generated for each of these extended corpora. The identified topics were inspected to identify which topics might be considered "off-task" for the corresponding collaborative activity (details of this are described in the Methods section). Thus, the topics were divided into two groups, namely *domain content relevant* and *irrelevant*.

Topic relevance The measure of group performance was operationalized as the amount of on-topic discussion. To develop a meaningful measure of relevant or "on-task" discussion, we

begin with the set of all topics, Q , constructed as described above. The topic score,

$$t_q(c_i) \quad (30)$$

gives the proportion of contribution c_i that covers topic $q \in Q$. These proportions sum to 1 for any contribution:

$$\sum_{q \in Q} t_q(c_i) = 1 \quad (31)$$

The set of all topics will be manually partitioned into two subsets, Q' and Q'' :

$$Q = Q' \cup Q''; Q' \cap Q'' = \{\emptyset\} \quad (32)$$

Q' represents those topics considered "relevant" or "on-task" for the corresponding collaborative activity, and Q'' consists of all other "off-task" topics (see the Method section). We can then construct a measure of the relative proportion of on-task material in each contribution by summing over the topic scores for topics in Q' :

$$T'(c_i) = \sum_{q \in Q'} t_q(c_i) \quad (33)$$

We can get a measure of the degree to which the entire group discussion was on- or off-task by averaging T' for all contributions across the entire discussion:

$$T' = \frac{1}{n} \sum_{i=1}^n T'(c_i) \quad (34)$$

Method

The GCA measures (as summarized in Table 2) were computed for each of three independent collaborative interaction datasets. The first was a Traditional Computer-Supported Collaborative Learning (CSCL) dataset. It is important to note that the Traditional CSCL dataset was the primary dataset used in the analyses. The second was a synchronous massive online course (SMOC) dataset called SMOC. The third was a

collaborative-learning and problem-solving dataset collected from a virtual internship game called “Land Science.” In the present research, the SMOC and Land Science datasets are used to address the external generalizability research question. The three datasets are described below.

Traditional CSCL dataset

Participants The participants were enrolled in an introductory-level psychology course taught in the fall semester of 2011 at a university in the American Southwest. Although 854 students participated in this course, some minor data loss occurred after removing outliers and those who failed to complete the outcome measures. The final sample consisted of 840 students. Females made up 64.3% of this final sample. In all, 50.5% of the sample identified as Caucasian, 22.2% as Hispanic/Latino, 15.4% as Asian American, 4.4% as African American, and less than 1% identified as either Native American or Pacific Islander.

Course details and procedure Students were told that they would be participating in an assignment that involved a collaborative discussion on personality disorders, as well as several quizzes. Students were told that their assignment was to log onto an online educational platform specific to the University at a specified time (Pennebaker, Gosling, & Ferrell, 2013). Students were also instructed that, prior to logging onto the educational platform, they should read certain assigned material on personality disorders.

After logging onto the system, students took a ten-item, multiple choice pretest quiz. This quiz asked students to apply their knowledge of personality disorders to various scenarios and to draw conclusions based on the nature of the disorders. After completing the quiz, they were randomly assigned to a chatroom with one to four classmates, also chosen at random (average group size was 4.59), and instructed to engage in a discussion of the assigned material. The group chat began as soon as someone typed the first message and lasted for exactly 20 min, when the chat window closed automatically. Then students took a second set of ten multiple-choice question posttest quiz. Each student contributed 154.0 words on average ($SD = 104.9$) in 19.5 sentences ($SD = 12.5$). As a group, discussions were about 714.8 words long ($SD = 235.7$) and 90.6 sentences long ($SD = 33.5$).

Group performance measure The group performance was operationally defined as the proportion of topic-relevant discussion during the collaborative interaction, as described in Eq. 34. As a reminder, the corpus used for the topic modeling was the same extended corpus (created using the seed method described earlier) used for creating the custom LSA spaces (Cai et al., 2011).

The topic modeling analysis revealed 20 topics, of which eight were determined to be relevant to the collaborative interaction task. Interjudge reliability was not used to determine the relevant topics. Instead, two approaches were used to determine the most relevant topics and to validate a topic relevance measure for group performance. The first was the frequency of the topics discussed across all the groups and individual students, wherein more frequently discussed topics were viewed as more important. Second, correlations between the topics and student learning gains were used to help validate the importance of the topic. Once the important topics were determined, an aggregate topic relevance score was computed by summing up the proportions for those topics (Eq. 33). The top ten words for each of the relevant topics are listed in the supplementary material.

SMOC dataset

Participants The participants were 1,713 students enrolled in an online introductory-level psychology course taught in the fall semester of 2014 at a university in the American Southwest. Throughout the course, students participated in a total of nine different computer-mediated collaborative interactions on various introductory psychology topics. This resulted in a total of 3,380 groups, with four to five students per group. However, 83 (2.46%) of the 3,380 chat groups were dropped because they contained only a single participant.

Course details and procedure The collaborative interactions took place in a large online introductory-level psychology course. The structure of the class followed a synchronous massive online course (SMOC) format. SMOCs are a variation of massive open online courses (MOOCs; Chauhan, 2015). MOOCs are open to the general public and typically free of charge. SMOCs are limited to a total of 10,000 students, including those enrolled at the university and across the world, and are available to all participants at a registration fee of \$550 (Chauhan, 2015).

The course involved live-streamed lectures that required students to log in at specific times. Once students were logged onto the university’s online educational platform, students were able to watch live lectures and instructional videos, take quizzes and exercises, and participate in collaborative discussion exercises. Students interacted in collaborative discussions via web chat with randomly selected classmates. Once put into groups, students were moved into a chat room and told they had exactly 10 min to discuss the assigned material (readings or videos). This 10-min session began at the moment of the first chat message. At the end of the discussion, students individually took a ten-item, multiple choice quiz that asked students to apply their knowledge of the assigned material to various scenarios and to draw conclusions.

Land Science dataset

Participants A total of 38 participants interacted in 19 collaborative problem-solving simulation games. Each game consisted of multiple rooms, and each room involved multiple chat sessions. There were a total of 630 distinct chat sessions. Of the 38 participants, $n = 29$ were student players, $n = 13$ were mentors, $n = 10$ were teachers, and $n = 1$ was a nonplayer character (NPC). For the purposes of detecting the social roles of players, only the players' and the mentors' chat was analyzed with the GCA. One of the rationales for exploring this dataset was to evaluate the generalizability of the GCA method across a range of different types of collaborative tasks. Specifically, unlike the collaborative-learning datasets described above, Land Science is a collaborative problem-solving environment.

Details and procedure Land Science is an interactive virtual urban-planning internship simulation with collaborative problem solving (Bagley & Shaffer, 2015; Shaffer, 2006; Shaffer & Graesser, 2010). The goal of the game is for students to think and act like STEM professionals. Players are assigned an in-game role as an intern with a land-planning firm in a virtual city, under the guidance of a mentor. During the game, players communicate with other members of their planning team, as well as with a mentor who sometimes role plays as a professional planning consultant. Players are deliberately given different instruction and resources; they must successfully combine skills within small teams in order to solve the collaborative problems.

Detecting social roles

The following analyses focus on addressing the main questions raised in the Overview of the Present Research, above. The analysis started with the Traditional CSCL dataset, which was immediately partitioned into training (84%) and testing (16%) datasets. Descriptive statistics for the GCA measures from the training data are presented in Table 3.

Table 3 Descriptive statistics for GCA measures

Measure	Minimum	Median	<i>M</i>	<i>SD</i>	Maximum
Participation	−0.26	−0.01	0.00	0.10	0.35
Social impact	0.00	0.18	0.18	0.05	0.43
Overall responsiveness	0.00	0.18	0.18	0.05	0.50
Internal cohesion	−0.06	0.18	0.18	0.09	0.58
Newness	0.00	0.48	0.78	1.25	18.09
Communication density	0.00	0.21	0.34	0.51	6.45

Mean (*M*). Standard deviation (*SD*).

The data were normalized and centered to prepare them for analysis. Specifically, the normalization procedure involved Winsorizing the data on the basis of each variable's upper and lower percentile. Density and pairwise scatter plots for the GCA variables are reported in the supplementary material. A cluster analysis approach was adopted to discover communication patterns associated with specific learner roles during collaborative interactions. Cluster analysis is a common data mining technique that involves identifying subgroups of data within the larger population who share similar patterns across a set of variables (Baker, 2010). Cluster analysis has been applied in previous studies of social roles (e.g., Lehmann-Willenbrock et al., 2016; Risser & Bottoms, 2014) and has proven useful in building an understanding of individuals' behaviors in many digital environments more broadly (del Valle & Duffy, 2007; Mirriahi, Liaqat, Dawson, & Gašević, 2016; Wise et al., 2012). Prior to clustering, multicollinearity was assessed through inflation factor (VIF) statistics and collinearity was assessed using Pearson correlations. The VIF results support the view that multicollinearity was not an issue with $VIF > 7$ (Fox & Weisberg, 2010). There was evidence of moderate collinearity between two variables, newness and communication density. However, further evaluation showed that collinearity did not impact the clustering results. For more details on collinearity and cluster tendency assessments, please see the [supplementary material](#).

In principle, any number of clusters can be derived from a dataset. So, the most important decision for any analyst when making use of cluster analysis is to determine the number of clusters that best characterizes the data. Several methods have been suggested in the literature for determining the optimal number of clusters (Han, Pei, & Kamber, 2012). A primary intuition behind these methods is that ideal clusterings involve compact, well-separated clusters, such that the total intracluster variation or total within-cluster sum of squares (wss) is minimized (Kaufman & Rousseeuw, 2005). In the present research, we used the NbClust R package, which provides 26 indices for determining the relevant number of clusters (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). It is beyond the scope of this article to specify each index, but they are described comprehensively in the original article of Charrad et al. An important advantage of NbClust is that researchers can simultaneously compute multiple indices and determine the number of clusters using a majority rule, wherein the proposed cluster size that has the best score across the majority of the 26 indices is taken to be optimal. Figure 3 reveals that the optimal number of clusters, according to the majority rule, was six for a *k*-means clustering. Note that two- and four-cluster solutions were also inspected and compared. In-depth coverage of those models and their evaluation may be found in the supplementary material.

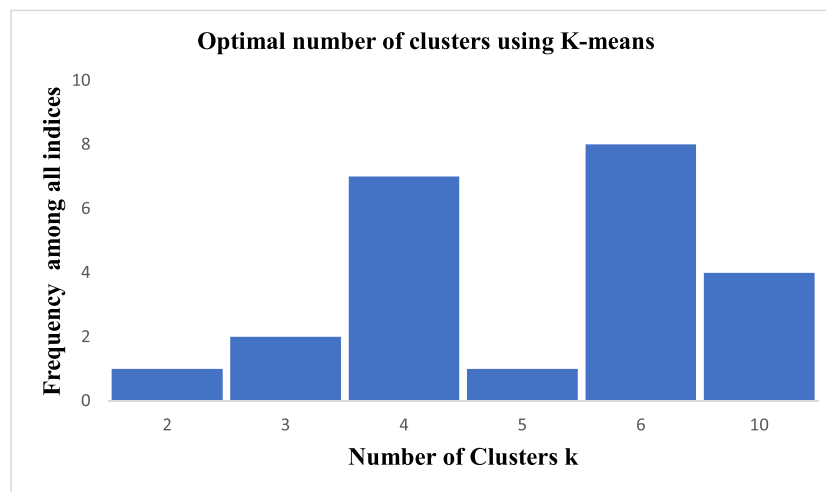


Fig. 3 Frequencies for recommended numbers of clusters using k -means, ranging from 2 to 10, based on the 26 criteria provided by the NbClust package. Here we see that eight of the 26 indices proposed six as the optimal number of clusters in the Traditional CSCL dataset.

Cluster analysis

K -means was used to group learners with similar GCA profiles into clusters. Investigation of the cluster centroids may shed light on whether the clusters are conceptually distinguishable. The centroids are representative of what may be considered typical, or average, of all entities in the cluster. With k -means, the centroids are in fact the means of the points in the cluster (although this is not necessarily true for other clustering methods). In the context of GCA profiles, we may interpret the centroids as behavior typical of a distinct style of interaction (i.e., roles). The centroids for the six-cluster k -means solution are presented in Fig. 4. It is worth noting that since the clustering was performed on normalized data, 0 in this figure represents the population average for each measure, whereas positive and negative values represent values above or below that average, respectively.

We see some interesting patterns across the six-cluster solution. Cluster 1 ($N = 143$) was characterized by learners who

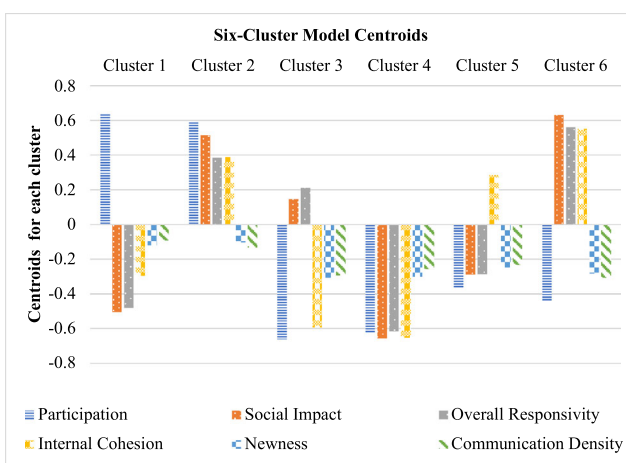


Fig. 4 Centroids for the six-cluster solution across the GCA variables.

had the highest participation and mid-range newness and communication density, but lower scores across all other measures. These individuals could be considered key members due to the sheer volume of their posting. However, the degrees to which these individuals were responding to others (i.e., overall responsiveness) and influencing the other group members' subsequent posts (i.e., social impact) are suggestive of behavior that was ineffective, or perhaps superficial, to the group interaction goals. Their discourse appears to have been more in response to themselves than to other group members, since they exhibit relatively higher internal cohesion than responsiveness. This relationship suggests that these learners may occasionally have had an influence similar to that of the Over-Riders described in Strijbos and De Laat's (2010) framework. The overall pattern of Cluster 1 highlights a theme in the literature, which suggests that high-volume members may not always be supportive in online interactions (Benamar, Balagué, & Ghassany, 2017; Nolker & Zhou, 2005). Nolker and Zhou raised the issue of separating high-contributing individuals into two classes, based on whether their conversational patterns were supportive of the collaboration or the prevailing social climate. These high-participating individuals in Cluster 1, who did not effectively contribute to productive group conversation as a whole, are labeled as *Chatterers* in the present research.

The learners in Cluster 2 ($N = 153$) were among the highest participators; they exhibit high social impact, responsiveness, and internal cohesion, but coupled with the lowest newness and communication density. Learners in these clusters were investing a high degree of effort in the collaborative discussion and displayed self-regulatory and social-regulatory skills. This pattern is labeled the *Drivers* in the present research.

Cluster 3 ($N = 88$) is characterized by learners who had the lowest participation. However, when they did contribute it appears to have built, at least minimally, on previously

contributed ideas and moved the collaborative discourse forward (i.e., moderately positive social impact and responsiveness). This cluster is labeled as *Followers*.

Cluster 4 ($N = 117$), labeled as *Lurkers*, is characterized by some of the lowest values across all GCA measures. Lurkers have been defined differently in the literature, ranging from nonparticipators to minimal participators (Nonnecke & Preece, 2000; Preece, Nonnecke, & Andrews, 2004). The distinction between a Ghost and a Lurker is not clear, and the terms appear to be interchangeable, although Strijbos and De Laat (2010) do make a distinction based on group size. There were two reasons that motivated us to prefer the term *Lurker*, rather than *Ghost*, in the present research. First, the GCA methodology would not be able to detect an individual who did not participate at all (because there would not be a log file for those students), which suggests that the learners in these clusters did contribute at least minimally. Second, past research has labeled the Ghost and Lurker roles predominantly on the basis of the amount of contribution that a student makes. However, the GCA captures participation as well as the sociocognitive characteristics of those contributions. Again, since these measures are normalized, the very low values for this cluster centroid does not suggest that these students had no social impact or were completely unresponsive to others. Rather, it suggests that these students expressed far less, as compared to the population average. Lurking behavior sometimes involves some level of engagement, but at other times little engagement, so it is associated with both positive and negative outcomes in the literature (Preece et al., 2004). Therefore, *Lurker* appeared to be the most appropriate label for Cluster 4.

Learners occupying Cluster 5 ($N = 91$) exhibited high internal cohesion but low scores on all the other GCA measures. This cluster is labeled as *Socially Detached*, because the pattern appears to capture students who were not productively engaged with their collaborative peers, but instead focused solely on themselves and their own narrative.

In Cluster 6 ($N = 126$) we see learners with low participation, but when they did contribute, they attended to other learners' contributions and provided meaningful information that furthered the discussion (i.e., high internal cohesion, overall responsiveness, and social impact). It is interesting to note that these students were not among the highest participators, but their discourse signaled a social positioning that was conducive to a productive exchange within the collaborative interaction. This pattern is suggestive of a student who is engaged in the collaborative interaction but takes a more thoughtful and deliberative stance than do the *Drivers*. As such, we refer to this cluster as *Influential Actors* in this research. Overall, the six-cluster model appears, at least upon an initial visual inspection, to produce theoretically meaningful participant roles. We then proceeded to evaluate the quality and validity of this model.

Clustering evaluation and validation

The literature has proposed several cluster validation indexes that quantify the quality of a clustering (Hennig, Meila, Murtagh, & Rocci, 2015). In principle, these measures provide a fair comparison of clustering and aid researchers in determining whether a particular clustering of data is better than an alternative (Taniar, 2006). Three main types of cluster validation measures and approaches are available: internal, stability, and external. *Internal* criteria evaluate the extent to which the clustering “fits” the dataset based on the actual data used for clustering. In the present research, two commonly reported internal validity measures (Silhouette and Dunn's index) were explored, using the R package *clValid* (Brock, Pihur, Datta, & Datta, 2008). Silhouette analysis measures how well an observation is clustered and it estimates the average distance between clusters (Rousseeuw, 1987). Silhouette widths indicate how discriminating the candidate clusters are by providing values that range from a low of -1 , indicating that observations are likely placed in the wrong cluster to 1 , indicating that the clusters perfectly separate the data and no better alternative clustering can be found. The average silhouette (AS) for the six-cluster model was positive ($AS = .31$), indicating the students in a cluster had higher similarity to other students in their own cluster than to students in any other cluster. Dunn's (1974) index (D) evaluates the quality of clusters by computing a ratio between the intercluster distance (i.e., the separation between clusters) and the intracluster diameter (i.e., the within-cluster compactness). Larger values of D suggest good clusters, and a D larger than 1 indicates compact separated clusters (Dunn, 1974). Dunn's index for the six-cluster model was $D = .5$, indicating that this clustering had moderate compactness.

Stability is another important aspect of cluster validity. A clustering may be said to be stable if its clusters remain intact (i.e., not disappear easily) when the dataset is changed in a nonessential way (Hennig, 2007). Although there may be many different conceptions of what constitutes a “nonessential change” to a dataset, the leave-one-column out method is commonly applied. The stability measures calculated in this way compare the results from clustering based on the complete dataset to clusterings based on removing each column, one at a time (Brock et al., 2008; Datta & Datta, 2003). In the present context, this corresponds to the removal of one of the GCA variables at a time. The stability measures are the average proportion of nonoverlap (APN), the average distance (AD), the average distance between means (ADM), and the figure of merit (FOM). Each of these measures was calculated for each reduced dataset (produced by dropping one column), and their average was taken as the measure for the dataset as a whole. The APN ranges from 0 to 1 , whereas the AD, ADM, and FOM all range from 0 to infinity. For each of these measures, smaller scores indicate a better, more stable

clustering. The stability scores for the six-cluster solution suggest that the clusters were quite stable across the four measures, with APN = .22, AD = 0.97, ADM = 0.31, and FOM = 0.37.

Cluster coherence

It is important to also evaluate the coherence of the clusters in terms of their underlying GCA variables. This can help to establish that the identified clusters do in fact represent distinct modes in the distribution of the GCA measures. Consequently, the six-cluster model was further evaluated to determine whether learners in the cluster groups significantly differed from each other on the six GCA variables. The multivariate skewness and kurtosis were investigated using the R package MVN (Korkmaz, Goksuluk, & Zararsiz, 2015), which produces the chi-square Q–Q plot (see the [supplementary material](#)), and a Henze–Zirkler (*HZ*) test statistic, which assesses whether the dataset follows an expected multivariate normal distribution. The results indicated the GCA variables did not follow a normal distribution, $HZ = 5.06, p < .05$. Therefore, a permutational multivariate analysis of variance (MANOVA; or a nonparametric MANOVA) was to evaluate the between-cluster GCA variable means. The permutational MANOVA, implemented in the *Adonis* routine of the VEGAN package in R (Oksanen et al., 2016), is a robust alternative to both the traditional parametric MANOVA and to ordination methods for describing how variation is attributed to different experimental treatments or, in this case, cluster partitions (Anderson, 2001). The *Adonis* test showed a significant main effect of the clusters, $F(5, 712) = 350.86, p < .001$. These results support the models' formation and ability to organize learners on the basis of differences in their collaborative communication profiles.

The analyses then proceeded through ANOVAs, followed by Tukey's post hoc comparisons to identify significant differences in the participants' scores on the six GCA variables between the clusters. Levene's test of equality of error

variances was violated for all the GCA variables, so a more stringent alpha level ($p < .01$) was used when identifying significant differences for these variables (Tabachnick & Fidell, 2007, p. 86). The ANOVA main effect *F* values, along with the means and standard deviations for the GCA variables across each cluster, are reported in Table 4 for the six-cluster model. The ANOVA revealed significant differences among clusters for all six of the GCA variables at the $p < .0001$ level for the six-cluster model. Tukey's HSD post hoc comparisons for the six-cluster model are shown in Table 5, where we can see that the observed differences in GCA profiles across the clusters were, for the majority, significantly distinct in both models.

Model generalizability

Internal generalizability When performing unsupervised cluster analyses, it is important to know whether the cluster results generalize (e.g., Research Question 2a). In the present research, a bootstrapping and replication methodology was adopted to see whether the observed clusters would generalize meaningfully to unseen data (Dalton, Ballarin, & Brun, 2009; Everitt, Landau, Leese, & Stahl, 2011). First, the internal generalizability was evaluated for the six-cluster model from the Traditional CSCL dataset. Specifically, a bootstrapping approach was used to assess the "prediction strength" of the training data. The prediction strength measure assesses how many groups can be predicted from the data, and how well (Tibshirani & Walther, 2005). Following the prediction strength assessment, a replication model was used to evaluate whether the training data cluster centroids can predict the ones in the testing data. If the six-cluster structure found using *k*-means clustering is appropriate for the Traditional CSCL data, then the prediction for the test dataset, and a clustering solution created independently for the test dataset, should match closely.

The prediction strength of the training data was explored using the *clusterboot* function in the R package *fpc* (Hennig, 2015). This approach uses a bootstrap

Table 4 Six-cluster model means and standard deviations for the six GCA variables

GCA Measures	Cluster 1: Chatters <i>n</i> = 143 <i>M</i> (<i>SD</i>)	Cluster 2: Driver <i>n</i> = 153 <i>M</i> (<i>SD</i>)	Cluster 3: Follower <i>n</i> = 88 <i>M</i> (<i>SD</i>)	Cluster 4: Lurker <i>n</i> = 117 <i>M</i> (<i>SD</i>)	Cluster 5: Detached <i>n</i> = 91 <i>M</i> (<i>SD</i>)	Cluster 6: Influential Actor <i>n</i> = 126 <i>M</i> (<i>SD</i>)	<i>F</i> value
Participation	0.64 (0.23)	0.60 (0.24)	− 0.66 (0.28)	− 0.63 (0.27)	− 0.37 (0.36)	− 0.44 (0.32)	285.70***
Social impact	− 0.50 (0.31)	0.51 (0.33)	0.15 (0.47)	− 0.66 (0.23)	− 0.29 (0.39)	0.63 (0.25)	200.50***
Overall responsivity	− 0.48 (0.31)	0.38 (0.38)	0.21 (0.46)	− 0.61 (0.24)	− 0.28 (0.36)	0.56 (0.28)	157.70***
Internal cohesion	− 0.30 (0.37)	0.39 (0.31)	− 0.59 (0.21)	− 0.65 (0.17)	0.29 (0.31)	0.55 (0.23)	210.30***
Newness	− 0.12 (0.14)	− 0.10 (0.14)	− 0.31 (0.14)	− 0.30 (0.13)	− 0.25 (0.15)	− 0.28 (0.12)	15.83***
Communication density	− 0.09 (0.14)	− 0.13 (0.16)	− 0.29 (0.15)	− 0.26 (0.15)	− 0.23 (0.16)	− 0.31 (0.12)	15.01***

ANOVA $df = 5, 712$; *** $p < .0001$.

Table 5 Tukey HSD p values for the pairwise comparisons for the GCA measures across the six-cluster solution

Cluster Comparison	GCA Variables					
	Participation	Social Impact	Overall Responsivity	Internal Cohesion	Newness	Communication Density
2 vs. 1	$p = .04$	$p < .001$	$p < .001$	$p < .001$	$p = .83$	$p = .06$
3 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
4 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
5 vs. 1	$p < .001$	$p = .008$	$p = .05$	$p < .001$	$p < .001$	$p < .001$
6 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
3 vs. 2	$p < .001$	$p < .001$	$p = .66$	$p < .001$	$p < .001$	$p < .01$
4 vs. 2	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .01$
5 vs. 2	$p < .001$	$p < .001$	$p < .001$	$p = .58$	$p < .05$	$p < .001$
6 vs. 2	$p < .001$	$p = .07$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
4 vs. 3	$p = .93$	$p < .001$	$p < .001$	$p = .99$	$p = 1.00$	$p = .99$
5 vs. 3	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = .56$	$p = .50$
6 vs. 3	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = .99$	$p = 1.00$
5 vs. 4	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = .61$	$p = .78$
6 vs. 4	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = 1.00$	$p = .98$
6 vs. 5	$p = .99$	$p < .001$	$p < .001$	$p < .001$	$p = .72$	$p = .37$

resampling scheme to evaluate the prediction strength of a given cluster. The algorithm uses the *Jaccard coefficient*, a similarity measure between sets. The Jaccard similarity between two sets Y and X is the ratio of the number of elements in the intersection of Y and X over the number of elements in the union of Y and X . The cluster prediction strength and stability of each cluster in the original six-cluster model is the mean value of its Jaccard coefficient over all the bootstrap iterations. As a rule of thumb, clusters with a value less than 0.6 should be considered unstable. Values between 0.6 and 0.75 indicate that the cluster is measuring a pattern in the data, but there is not high certainty about which points should be clustered

together. Clusters with values above about 0.85 can be considered highly stable and have high prediction strength (Zumel, Mount, & Porzak, 2014). The prediction strength of the Traditional CSCL training data was evaluated using 100 bootstrap resampling iterations.

The final cluster pattern produced by the 100 bootstrap resampling iterations for the six-cluster model are reported in Fig. 5. As can be seen in this figure, the observed pattern was identical to the original six-cluster model, albeit with a different ordering of the clusters. The ordering of clusters in the k -means algorithm is arbitrary, so the pattern of the GCA variables within each cluster was of most importance. The Jaccard similarity values showed very strong prediction for

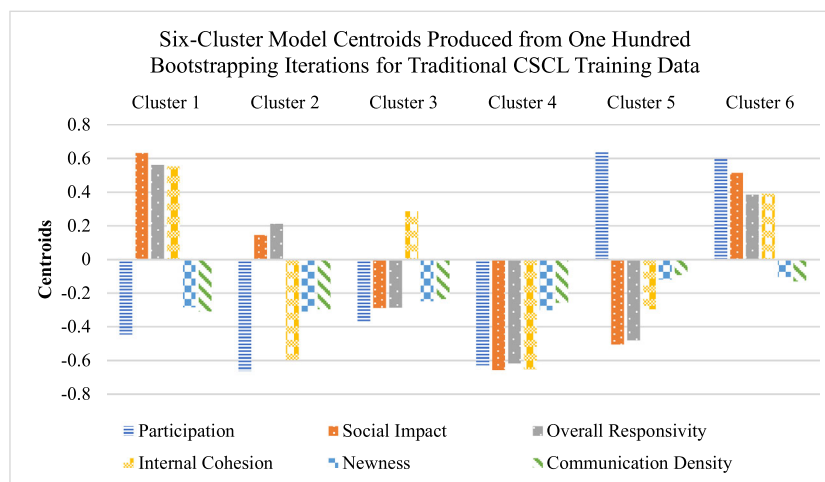


Fig. 5 Final six-cluster pattern produced by the 100 bootstrap resampling iterations of the Traditional CSCL training data, which was identical to the original k -means six-cluster model pattern depicted in Fig. 4.

all six clusters, with .96, .95, .91, .96, .91, and .96 for Clusters 1–6, respectively.

The next analyses focused on evaluating the generalizability of the observed clusters in the training data to the testing data. First, six-cluster *k*-means analyses were performed on the held-out Traditional CSCL test data ($N = 136$). Descriptive statistics for the test data GCA variables are reported in the supplementary material. The centroids for the six-cluster *k*-means solution for the Traditional CSCL test data are illustrated in Fig. 6. The observed pattern of the six-cluster solution for the testing data appears, at least visually, to be similar to the one observed for the training data.

Next, we focused on quantifying the observed overlap between the testing and training cluster analyses. Specifically, the cluster centers from the training dataset were used to predict the clusters in the test data for the six-cluster model. This analysis was performed using the `cl_predict` function in the R `clue` package (Hornik & Böhm, 2016). Cross-tabulations of the predicted and actual cluster assignments for the Traditional CSCL testing dataset are reported in Table 6. The rows in the table correspond to the clusters specified by the *k*-means clustering on the testing data, and the columns correspond to the predicted cluster memberships from the training data. In a perfect prediction, large values would lie along the diagonal, with zeroes off the diagonal; this would indicate that all samples that belonged to Cluster 1 had been predicted by the training data as belonging to Cluster 1, and so forth. The form of this table can give us considerable insight into which clusters were reliably predicted. It can also show which groups are likely to be confused and which types of misclassification are more common than others. However, in this case we observed almost perfect prediction of the six-cluster model, with few exceptions.

Two measures were used to evaluate the predictive accuracy of the six-cluster model on the Traditional CSCL training clusters: the adjusted Rand index (ARI) and a measure of

Table 6 Cross-tabulations of the predicted and actual cluster assignments for the six-cluster model on the Traditional CSCL testing dataset

Testing Clusters	Training Predicted Clusters					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	32	0	0	0	0	0
Cluster 2	2	29	0	0	0	0
Cluster 3	0	0	15	2	1	0
Cluster 4	0	0	0	18	0	0
Cluster 5	4	0	0	1	13	0
Cluster 6	0	0	0	0	0	19

effect size (Cramer's *V*) for the cluster cross-tabulation. ARI computes the proportion of the total of $\binom{n}{2}$ object pairs that agree—that is, that are either (i) in the same cluster according to Partition 1 and the same cluster according to Partition 2 or (ii) in different clusters according to Partition 1 and in different clusters according to Partition 2. The ARI addresses some of the limitations of the original Rand index by providing a conservative measure that penalizes for any randomness in the overlap (Hubert & Arabie, 1985). The ARI was calculated between: (a) the test data clustering membership and (b) the predicted cluster membership given by the training data. The predictive accuracy of the training data is considered good if it is highly similar to the actual testing data cluster membership. The degree of association between the membership assignments of the predicted and actual cluster solutions was $ARI = .84$ for the six-cluster model. ARI values range from 0 to 1, with higher index values indicating more agreement between sets. The measure of effect size for the cross-tabulation revealed Cramer $V = .92$, which is considered very strong association (Kotrlík, Williams, & Jabor, 2011). Given these

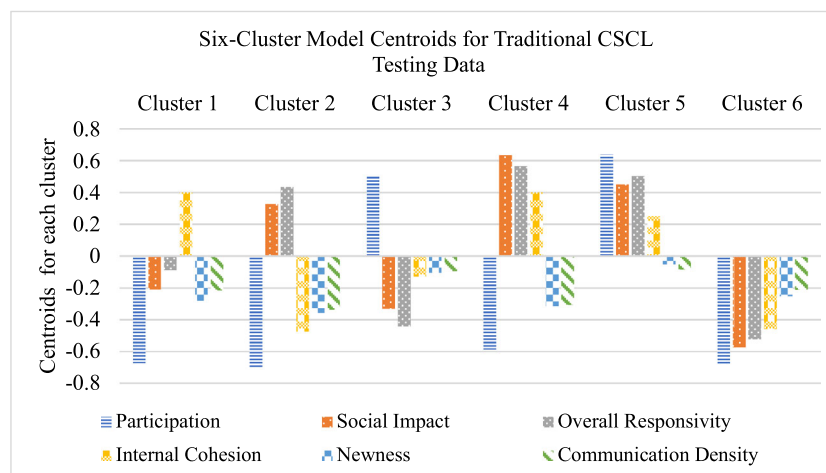


Fig. 6 Traditional CSCL testing data centroids for the six-cluster solution across the GCA variables.

results, the six-cluster solution was judged to be robust and well-supported by the data.

A similar replication approach was adopted to evaluate the generalizability within the SMOC and Land Science datasets. Descriptive statistics for the GCA measures in the SMOC training ($N = 9,463$)/testing ($N = 2,378$) and Land Science training ($N = 2,837$)/testing ($N = 695$) datasets are presented in Table 7. First, a six-cluster model was constructed on the SMOC and Land Science training datasets. The patterns of the six-cluster models are depicted in Fig. 7 for the SMOC training dataset, and in Fig. 8 for the Land Science training dataset.

The analysis proceeded by evaluating the internal generalizability for the SMOC and Land Science datasets separately. This analysis was performed by using the cluster centroids from the SMOC and Land Science training datasets to predict the clusters in the test data for the six-cluster model. These analyses were also performed using the `cl_predict` function in the R `clue` package (Hornik & Böhm, 2016). Cross-tabulations of the predicted and actual cluster assignments for the SMOC and Land Science testing dataset are reported in Tables 8 and 9, respectively. We see from these tables that there appears to be good agreement for the predicted cluster assignments in the six-cluster models. We can quantify the agreement using ARI and Cramer V , provided by the `flexclust` package. For the SMOC and Land Science datasets, $ARI = .90$ and $ARI = .86$, respectively. Again, the ARI values range from 0 to 1, with higher index values indicating more agreement between sets. This suggests that the six-cluster model exhibited slightly higher predictive agreement between the training and testing data cluster assignments for the SMOC than for the

Land Science dataset. However, both the SMOC and Land Science datasets had high effect sizes, with Cramer $V = .95$ and $.92$, respectively. Taken together, the six-cluster solutions were judged to be supported by both the SMOC collaborative interaction data and the Land Science collaborative problem-solving data, with the six-cluster model having only minimally better internal generalizability.

External generalizability The practice of predictive modeling defines the process of developing a model in a way that we can understand and quantify the model's prediction accuracy on future, yet-to-be-seen data (Kuhn & Johnson, 2013). The previous analyses provided confidence in the six-cluster models' ability to generalize to unseen data within the same dataset. However, the ultimate goal is to evaluate how well the identified student roles (i.e., clusters) are representative of interaction patterns across various types of collaborative interactions. This step is critical, because the robustness and accuracy of the models across datasets will determine the usefulness of the GCA for broader research applications. Thus, the next analyses assess the generalizability of these clusters across the three collaborative interaction datasets (i.e., Research Question 2b). Specifically, the clusters centers from each dataset were used to predict the clusters in the other training datasets, wherein all possible combinations were evaluated. Again, two measures were used to evaluate the predictive accuracy of the clusters, ARI and a measure of effect size, Cramer V , for their cross-tabulation. Table 10 shows the ARI and Cramer V results for the computed cross-tabulation evaluations of the six-cluster models. The columns in Table 10 correspond to the predictor dataset, whereas the rows correspond to the predicted dataset.

Table 7 Descriptive statistics for GCA measures in the SMOC and Land Science training and testing datasets

Measure	Min		Med		M		SD		Max	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
SMOC										
Participation	-0.44	-0.49	0.00	0.00	0.00	0.00	0.11	0.11	0.45	0.42
Social impact	-0.14	-0.05	0.15	0.15	0.16	0.16	0.10	0.10	1.00	1.00
Overall responsiveness	-0.30	-0.04	0.15	0.15	0.16	0.16	0.11	0.11	1.00	1.00
Internal cohesion	-0.43	-0.17	0.12	0.12	0.13	0.14	0.12	0.12	1.00	1.00
Newness	0.00	0.00	0.65	0.65	0.84	0.83	0.83	0.76	17.39	7.73
Communication density	0.00	0.00	0.19	0.19	0.26	0.26	0.30	0.26	10.56	3.32
Land Science										
Participation	-0.50	-0.49	-0.01	-0.03	0.00	0.00	0.14	0.15	0.78	0.49
Social impact	-0.10	-0.05	0.12	0.12	0.13	0.12	0.09	0.08	0.90	0.74
Overall responsiveness	-0.12	-0.04	0.11	0.11	0.13	0.12	0.10	0.09	1.00	1.00
Internal cohesion	-0.21	-0.17	0.11	0.11	0.13	0.12	0.13	0.12	1.00	1.00
Newness	0.00	0.00	0.60	0.59	1.10	1.11	2.33	2.15	70.27	27.39
Communication density	0.00	0.00	0.18	0.18	0.38	0.36	0.94	0.72	31.27	10.45

Mean (M). Standard deviation (SD).

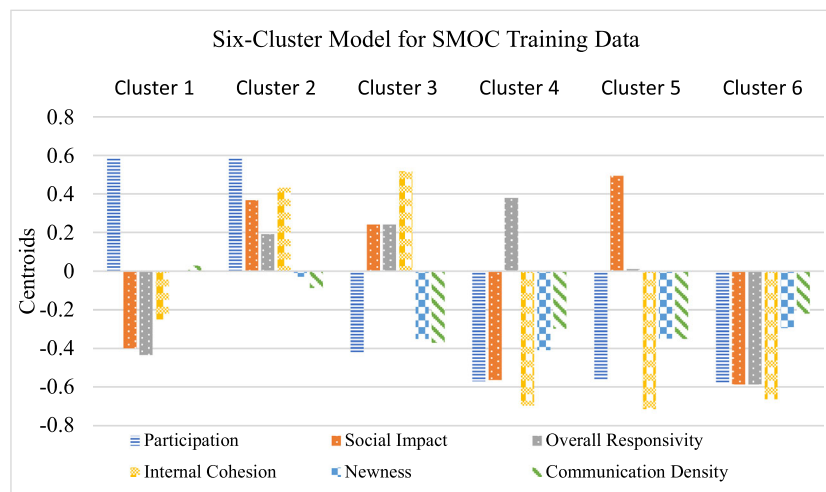


Fig. 7 SMOC training data centroids for the six-cluster solution across the GCA variables.

The first insight to take away from Table 10 is that the predictive accuracy (ARI) is lower for all datasets than in the previously reported internal generalization evaluations. This overall drop in predictive accuracy is to be expected with evaluating external data. Although the accuracy is lower than in the internal evaluations, the ARI results are still quite high for the majority of the predictions. Specifically, we see that the SMOC dataset has the lowest agreement predicting clusters in the Traditional CSCL and Land Science. However, Land Science had the highest agreement for predicting the Traditional CSCL data, and was on a par with the Traditional CSCL when predicting the SMOC dataset.

Student roles and learning Unlike the internal criteria explored in the section above, external criteria are independent of the ways clusters are obtained. External cluster validation can be explored by either comparing the cluster solutions to some “known” categories or by comparing them to meaningful external variables—that is, variables not used in the cluster

analysis (Antonenko, Toy, & Niederhauser, 2012). Furthermore, the practical impact of the identified social roles may be felt at multiple of levels of granularity, and we therefore must test for their impact at multiple levels. In the present research, the usefulness of identifying learners’ roles in collaborative learning was explored through two analyses of the data: (a) the influence of student roles on individual students’ performance and (b) the influence of student roles on overall group performance (Research Questions 3a and 3b).

The multilevel investigation conducted in the present research also addressed a frequently noted limitation found in collaborative-learning research. CSCL researchers encounter issues regarding the differing units of analysis in their datasets (Janssen, Erkens, Kirschner, & Kanselaar, 2011). That is, collaborative interactions can be analyzed at the level of the group, the individual student, and of each student-student interaction. For example, in the present research, some variables of interest were measured at the individual learner and interaction levels (e.g., student learning gains, participation,

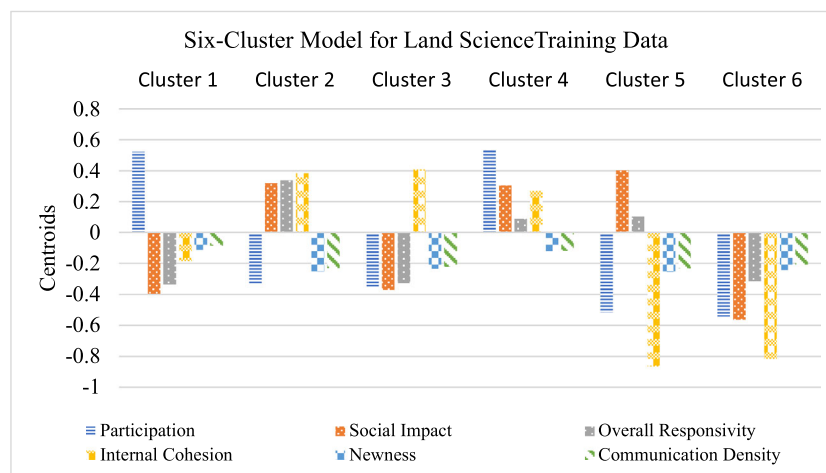


Fig. 8 Land Science training data centroids for the six-cluster solution across the GCA variables.

Table 8 Cross-tabulations of the six-cluster model predicted and actual cluster assignments for the SMOC testing dataset

Testing Clusters	Predicted Clusters					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	517	17	4	0	1	15
Cluster 2	0	469	14	0	0	0
Cluster 3	0	5	475	1	0	10
Cluster 4	1	0	1	208	0	4
Cluster 5	0	0	6	6	198	0
Cluster 6	1	0	0	3	7	415

internal cohesion, social impact, overall responsiveness, newness, communication density, and social roles identified by the cluster analysis), whereas other variables were measured at the group level (e.g., group diversity, group composition, and group performance). Several researchers have emphasized the need to conduct more rigorous, multilevel analyses (Cress, 2008; De Wever, Van Keer, Schellens, & Valcke, 2007; Stahl, 2005; Suthers, 2006). However, collaborative-learning studies have usually focused on only one of these levels (Stahl, 2013). As a result, little consideration has been given to how these levels are connected, despite its being well-recognized that such connections are crucially important to both understanding and orchestrating learning in collaborative-learning environments (Stahl, 2013). To avoid this problem, a series of models were constructed to explore both the influence of group-level constructs on individual student-level learning, as well as individual student-level constructs on group performance.

A student-level performance score was obtained for each student by calculating their proportional learning gains, formulated as in (Hake, 1998):

$$\frac{(\%PostTest - \%PreTest)}{(100 - \%PreTest)} \quad (35)$$

Table 9 Cross-tabulations of the six-cluster model predicted and actual cluster assignments for the Land Science testing dataset

Testing Clusters	Predicted Clusters					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	137	0	0	0	1	1
Cluster 2	0	90	3	9	4	0
Cluster 3	1	12	81	0	0	0
Cluster 4	11	0	2	106	0	0
Cluster 5	0	0	0	0	98	0
Cluster 6	0	0	0	0	1	138

The correlations between learning gains and the six GCA variables in the Traditional CSCL dataset are reported in Table 11.

A mixed-effects modeling methodology was adopted for these analyses, due to the nested structure of the data (e.g., students within groups; Pinheiro & Bates, 2000). Mixed-effects models include a combination of fixed and random effects, and can be used to assess the influence of the fixed effects on dependent variables after accounting for the random effects. Multilevel modeling handles the hierarchical nesting, interdependency, and unit of analysis problems that are inherent in collaborative-learning data. This is the most appropriate technique for investigating data in CSCL environments (De Wever et al., 2007; Janssen et al., 2011). Table 12 provides an overview of the mixed-effects models used to explore such potential multilevel effects across the six-cluster solution.

In addition to constructing the fixed-effects models, null models with the random effects (the student nested in the group or the group) but no fixed effects were also constructed. A comparison of the null random-effects-only model with the fixed-effect models allowed us to determine whether social roles and communication patterns predicted student and group performance above and beyond the variance attributed to individual students or groups. The Akaike information criterion (AIC), log likelihood (LL), and likelihood ratio test were all used to evaluate the overall fits of the models. Additionally, the effect sizes for each model were estimated using a pseudo- R^2 method, as suggested by Nakagawa and Schielzeth (2013). For mixed-effects models, R^2 can be divided into two parts: marginal (R^2_m) and conditional (R^2_c). Marginal R^2 is associated with variance explained by fixed factors, whereas conditional R^2 can be interpreted as the variance explained by the entire model, namely random and fixed factors. Both the marginal and conditional parts convey unique and relevant information regarding the model fit and variance explained. The nlme package (Pinheiro et al., 2016) was used to perform all the required computations. All analyses were on the Traditional CSCL dataset, because it was the base corpus for the cluster analyses and had the most consistent individual and group performance measures.

Influence of student roles on individual student performance

To evaluate the effects of roles at the purely individual level, two linear mixed-effects models were compared: (a) Model 1 from Table 12, with learning gains as the dependent variable, social roles as independent variables, and student nested within group as the random effects, and (b) the null model, with random effects only and no fixed effects. The likelihood ratio tests indicated that the six-role model, with $\chi^2(5) = 11.55$, $p = .04$, $R^2_m = .02$, $R^2_c = .95$, yielded a significantly better fit than the null model. A number of conclusions can be drawn from these statistics. First, the roles in the six-cluster model were able to add significantly to the prediction of the learners'

Table 10 ARI and Cramer's *V* results for the cluster model computed cross-tabulation tables

Six-Cluster Model	W3 Training		SMOC Training		Land Science Training	
	ARI	Cramer <i>V</i>	ARI	Cramer <i>V</i>	ARI	Cramer <i>V</i>
W3 training data	.84	.92	.66	.89	.76	.86
SMOC training data	.70	.78	.90	.95	.69	.79
Land Science training data	.69	.83	.66	.78	.86	.92

performance, beyond the variance attributed to student and group membership. Second, social roles, individual participant, and group features explained about 95% of the predictable variance, with 2% of the variance being accounted for by the social roles.

The social roles that were predictive of individual student learning performance for the six-cluster model are presented in Table 13. The reference group was the Driver role, meaning that the learning gains for the other roles are compared against the Driver reference group. Four of the six roles exhibit significant differences in student learning gains, as compared to the Driver role. Here we see that learners who took on more socially responsible, collaborative roles, such as the Driver, performed significantly better than did students who occupied the less socially engaged roles, including Lurker and Chatterer. There was no significant difference between the performance of the Drivers and the Influential Actor and Socially Detached roles, suggesting that these are the more successful roles in terms of student learning gains.

It is important to note that the observed difference in learning gains across the social roles was not a result of the students simply being more prolific, because the Influential Actor and Socially Detached learners performed on a par with the Drivers but were among of the lower participators in the group. The profile for the Socially Detached learners showed mid-range values for responsivity and social impact, as compared to their internal cohesion scores. However, the Influential Actor profile illustrated that when these students did make contributions, they were very responsive to the other group

members (i.e., high overall responsivity), as well as being semantically connected with their own previous contributions (i.e., high internal cohesion). Furthermore, their contributions were seen as relevant by their peers (i.e., high social impact). These findings reflected a more substantive difference in social awareness and engagement for the Drivers and Influential Actors than for the Chatterers, beyond the surface-level mechanism of simply participating often. Taken together, these results show that the identified roles are externally valid—not just because of a significant relationship to the external measure of learning, but also because we can make theoretically meaningful predictions from the roles associated with characteristic behaviors.

Incorporating group-level measures

Two groups of models were constructed to assess the influence of group composition on group performance and individual student learning gains. The first set of models (i.e., Models 2 and 3 from Table 12) assessed the influence of group role diversity on student learning gains and group performance, respectively. The second set of models (i.e., Models 4–6 and 7–9 from Table 12) dove deeper, to explore the influence of group compositions, as measured by the proportional occurrence of each of the roles, on student learning gains and group performance, respectively. As a reminder, group performance was operationally defined as the amount of topic-relevant discussion during the collaborative interaction (Eq. 34).

Table 11 Correlations between learning and GCA variables in the traditional CSCL dataset

	Learning Gains	Participation	Social Impact	Overall Responsivity	Internal Cohesion	Newness
Participation	.10**					
Social impact	.10*	.07				
Overall responsivity	.10*	-.01	.69***			
Internal cohesion	.13***	.21***	.57***	.52***		
Newness	.06	.62***	.05	-.03	.11**	
Communication density	.04	.54***	-.11***	-.18***	-.05	.91***

*** $p < .001$. ** $p < .01$. * $p < .05$.

Table 12 Overview of mixed-effects models exploring learning across the six-cluster solution

Model Number	Dependent Variable	Level of Dependent Variable	Independent Variable	Level of Independent Variable	Random Variable(s)
1	Learning gains	Student	Social roles	Student	Student nested in group
2	Learning gains	Student	Role diversity	Group	Student nested in group
3	Performance	Group	Role diversity	Group	Group
4–6	Learning gains	Student	Proportional occurrence roles	Group	Student nested in group
7–9	Performance	Group	Proportional occurrence roles	Group	Group

The proportional occurrence (frequency) of each role within any group can be a helpful measure for determining group composition. For a group G , it can be formally defined as

$$\hat{p}_G(r) = \frac{\# \text{users of role } r \text{ in } G}{\text{size of } G}$$

Group compositions was operationalized using a measure of role diversity based on entropy. Entropy is a measure at the core of information theory quantifying the amount of “surprise” possible in a probability distribution. At the extremes, entropy ranges from values of 0 for distributions in which a single outcome is always the case [i.e., $P(X=x) = 1.0$], to a maximum value when the probability of all outcomes is equal (i.e., a uniform distribution). The entropy of roles in a group will then be 0 for groups in which all participants take on the same role, and greater for groups with a greater diversity of roles. Role diversity for a group is calculated as

$$H(G) = - \sum_{r \in \text{Roles}} \hat{p}_G(r) \cdot \log(\hat{p}_G(r))$$

Correlations between group performance, student learning gains, role diversity, and the proportional occurrences of each role are reported in Table 14. No relationship was observed between student learning gains and group performance, so this was not probed further. Quite small relationships were observed between role diversity ($M = 1.04$, $SD = .26$) and both

Table 13 Descriptive statistics for student learning gains across the six roles, and mixed-effects model coefficients predicting differences in individual student performance across clusters

Role	Six-Cluster Model			
	M	SD	β	SE
Driver	0.21	0.89	0.21**	0.07
Chatterer	0.02	0.88	-0.19*	0.10
Lurker	-0.11	0.79	-0.32**	0.11
Follower	-0.08	0.92	-0.29**	0.12
Socially Detached	0.03	0.83	-0.18	0.11
Influential Actor	0.09	0.84	-0.12	0.10

* $p < .05$. ** $p < .01$. *** $p < .001$. Mean (M). Standard deviation (SD). Fixed-effect coefficient (β). Standard error (SE).

student learning gains and group performance. However, when these relationships were further explored, the likelihood ratio tests indicated that the full diversity models for student learning gains and group performance did not yield a significantly better fit than the null model, with $\chi^2(1) = 0.39$, $p = .52$, $R^2_m = .001$, $R^2_c = .96$, and $\chi^2(1) = 0.26$, $p = .62$, $R^2_m = .002$, $R^2_c = .88$, respectively.

The second set of analyses involved a more fine-grained investigation of the influence of (the proportional occurrence of) positive and negative roles on student learning gains and group performance. Six linear mixed-effects models were constructed, where three of the models were student-level (i.e., Models 4–6 from Table 12), and the other three were group-level (i.e., Models 7–9 from Table 12). Particularly, we constructed a *productive-roles model*, with the proportional occurrence of Drivers, Influential Actors, and Socially Detached learners as the independent variable, and an *unproductive-roles model*, with the proportional occurrence of Chatterers, Followers, and Lurkers as the independent variable. Null models were constructed for both the student- and group-level analyses. For the six models below, the first three had student learning gains as the dependent variable, whereas the next three had group performance as the dependent variable.

For the student-level analyses, the likelihood ratio tests indicated that neither the productive-roles model nor the unproductive-roles model yielded a significantly better fit than the null model, with $\chi^2(3) = 2.62$, $p = .45$, $R^2_m = .004$, $R^2_c = .96$, and $\chi^2(3) = 2.75$, $p = .43$, $R^2_m = .004$, $R^2_c = .96$, respectively. When we combined this with the previous finding that social role did influence individual learning, this suggests that it is less important that productive roles be present in one’s group than that the individual be enacting a productive role.

For the group-level analysis, the likelihood ratio tests indicated that both the productive-roles and the unproductive-roles models yielded significantly better fits than the null model, with $\chi^2(3) = 23.62$, $p < .0001$, $R^2_m = .15$, $R^2_c = .90$, and $\chi^2(3) = 20.92$, $p < .001$, $R^2_m = .13$, $R^2_c = .89$, respectively. Several conclusions can be drawn from this model comparison. First, the proportional occurrences of both productive and unproductive roles were able to significantly improve predictions of group performance, above and beyond the variance

Table 14 Correlations between student learning gains, group performance, role diversity, and the proportional occurrence of six roles

Measure	Student Level	Group Level Measures						
	Learning Gains	Group Performance	Diversity	Prop. Chatterers	Prop. Driver	Prop. Follower	Prop. Lurker	Prop. Socially Detached
Group Performance	.00							
Diversity	-.02	-.03						
Prop. Chatterers	-.03	-.28***	.03					
Prop. Driver	.03	.28***	-.12***	-.77***				
Prop. Follower	-.01	.02	.12**	-.31***	.29***			
Prop. Lurker	-.05	-.28***	-.04	.47***	-.49***	-.46***		
Prop. Socially Detached	-.01	-.13***	.23***	.16***	-.43***	-.29***	.07	
Prop. Influential Actor	.05	.32***	-.16***	-.47***	.28***	-.11**	-.52***	-.37***

*** $p < .001$. ** $p < .01$. * $p < .05$.

attributed to the group. Second, for all models, the proportional occurrence of different social roles, combined with group features, explained about 89% of the predictable variance in group performance, with 28% of the variance being accounted for by the proportional occurrence of different social roles. Table 15 shows the social roles that were predictive of group performance for both the productive-roles and unproductive-roles models.

As is shown in Table 15, groups with greater proportions of learners who took on more socially responsible, collaborative roles (namely, Drivers and Influential Actors) performed significantly better than groups with greater proportions of less socially engaged roles (Lurkers and Chatterers). These findings mirror the pattern that we observed for individual student learning and social roles.

Discussion

Detecting roles

In the Detecting Social Roles section, we explored the extent to which the characteristics of collaborative interaction discourse, as captured by the GCA, diagnostically reveal the social roles students occupy, and if the observed patterns are

robust and generalize. The GCA was applied to two large, collaborative-learning datasets, and one collaborative problem-solving dataset (learner $N = 2,429$, group $N = 3,598$). Participants were then clustered on the basis of their profiles across the GCA measures. The cluster analyses identified roles that have distinct patterns in behavioral engagement style (i.e., active or passive, leading or following), contribution characteristics (i.e., providing new information or echoing given material), and social orientation. The six-cluster model revealed the following roles: Drivers, Influential Actors, Socially Detached learners, Chatterers, Followers, and Lurkers.

The findings present some methodological, conceptual and practical implications for the group interaction research, educational data mining and learning analytics communities. The GCA represents a novel methodological contribution, capable of identifying distinct patterns of interaction representative of the social roles students occupy in collaborative interactions. The natural language metrics that make up the GCA provide a mechanism to operationalize such roles, and provide a view on how they are constructed and maintained through the sociocognitive processes within an interaction. We expect the GCA to provide a more objective, domain independent, and deeper exploration of the micro-level inter- and intra-personal patterns associated with social roles. Moreover, as

Table 15 Descriptive statistics for group performance across the six roles, and mixed-effects model coefficients for predicting the influences of productive and unproductive roles on group performance

Role	Productive Roles Model				Role	Unproductive Roles Model			
	<i>M</i>	<i>SD</i>	β	<i>SE</i>		<i>M</i>	<i>SD</i>	β	<i>SE</i>
Prop. Driver	0.27	1.05	1.15**	0.41	Prop. Chatterers	-0.27	0.92	-1.05*	0.46
Prop. Socially Detached	-0.18	0.79	0.42	0.52	Prop. Follower	0.03	0.94	-1.02*	0.55
Prop. Influential Actor	0.37	1.04	1.27**	0.39	Prop. Lurker	-0.32	0.94	-1.42*	0.52

$N = 148$. * $p < .06$. ** $p < .01$. Mean (*M*). Standard deviation (*SD*). Fixed-effect coefficient (β). Standard error (*SE*).

the methodology is readily automated, substantially larger corpora can be analyzed with the GCA than is practical when human judgments are required to annotate the data.

The identified social roles (i.e., clusters) underwent stringent evaluation, validation, and generalization assessments. The bootstrapping and replication analyses illustrated that the roles generalize both within and across different collaborative interaction datasets, indicating that these roles are robust constructs across different experimental contexts. Given the extent of these evaluations, we feel that the roles identified can be considered as robust and stable constructs in the space of small group interactions, and that the GCA measures capture the critical sociocognitive processes necessary for identifying such roles.

The present research has built upon the framework of Strijbos and De Laat (2010) by adding several new dimensions of interaction. Interestingly, the GCA revealed roles that do not entirely overlap with those observed in Strijbos and De Laat's. In this respect, we have been able to build upon their results and to provide insights beyond what their framework revealed. The identification of these additional roles may serve as a useful conceptual addition for future research focusing on the social roles within multiparty communication. For instance, only one of Strijbos and De Laat's roles, the Over-Rider, appeared similar to a group (i.e., the Chatterers in the present research) in the six-cluster model for the Traditional CSCL dataset. However, the other roles did not appear to align with the labels suggested Strijbos and De Laat's framework. This is likely due to the fact that the GCA includes more and different dimensions than are represented in the previous framework.

Roles and learning

In the Student Roles and Learning section above, we investigated the practical value of the identified roles, and whether they were meaningfully related to student learning gains and group performance. Overall, the results suggest that the roles that learners occupy influences their learning, and that the presence of specific roles within a group can be either more or less beneficial for the collaborative outcome. Furthermore, we established the connection between the individual-level and group-level outcomes is affected by the same productive or unproductive roles. Taken together, these discoveries show that not only are the identified roles related to learning and to collaborative success, but that this relationship is theoretically meaningful, which provides external validity.

This analysis yielded two important contributions to the collaborative-learning literature. Firstly, the multilevel mixed-effects models applied in this chapter are rarely applied in CSCL research; however, they are the most appropriate statistical analysis for this nested structure data CSCL data

(De Wever et al., 2007; Janssen et al., 2011; Pinheiro & Bates, 2000). Furthermore, these models impose a very stringent test of the influence of roles on group and individual performance by controlling for the variance associated with each participant and group. As such, the use of mixed-effects models provides confidence in the robustness of the findings. Second, the multilevel investigation addressed a frequently noted limitation found in collaborative-learning research. As Kapur, Voiklis, and Kinzer (2011, p. 19) wrote:

It is worth reiterating that these methods should not be used in isolation, but as part of a larger, multiple grain size analytical program. At each grain size, findings should potentially inform and be informed by findings from analysis at other grain sizes—an analytical approach that is commensurable with the multiple levels (individual, group) at which the phenomenon unfolds. Only then can these methods and measures play an instrumental role in the building and testing of a process-oriented theory of problem solving and learning.

Some of the most noteworthy of the present discoveries concern the influence of roles on student learning and group performance. For the individual student learning models, we saw that socially engaged roles, like Driver, significantly outperformed less participatory roles, like Lurkers. This finding might be expected. However, other findings emerged that were less intuitive. For instance, we found that Influential Actor and Socially Detached learners performed comparably well with the Drivers (although not quite as high), but were among of the lower participators in the group. This suggests the difference in learning gains across the social roles is not simply a result of the students being more prolific. Clearly, engagement with or mastery of the material can be manifested not only through greater quantity, but also greater quality, of participation. The Influential Actors were highly responsive and had high social impact and internal cohesion, but lower scores for newness and communication density. However, the most defining feature of the Socially Detached learners was their high internal cohesion because they exhibited relatively mediocre scores across the other GCA measures. Something interesting starts to emerge when these profiles are juxtaposed with the Chatterers. The Chatterers were the highest participators, but had lower learning gains, responsiveness to peers, social impact, and mediocre internal cohesion than those in other groups. Together, this highlights that participating a lot is far less important than is the nature of that participation (i.e., the intra- and interpersonal dynamics as captured by the internal cohesion, responsiveness, and social impact measures). That is, the quality of conversation, more than the quantity, appears to be the key element in the success for both groups and individuals.

The influence of the roles on group performance was also investigated. We started by looking at the influence of the overall diversity of roles on group performance. Here, we were interested in seeing if groups that are comprised of, for example, six different roles performed better than those that were comprised of all Influential Actors. This was motivated by the group interaction literature, which suggests that diversity can be a major contributor to the successfulness of collaborative interactions. The findings for diversity in the literature have explored several different types of diversity, including personality, prior knowledge, gender, and other individual traits (Barron, 2003; Fuchs, Fuchs, Hamlett, & Karns, 1998). These analyses did not reveal any significant influence of role diversity on student or group performance, suggesting, perhaps, that diversity in roles is not an important type of diversity. It is important to note that the attributes explored in previous work have primarily focused on “what” students bring to the group, rather than how students engage in the group. This could possibly explain why diversity in roles was not as important as the other types of diversity in the literature.

We then dove deeper into an investigation of group composition as given by the proportional occurrence of each role. The findings here were considerably more promising, and largely mirrored those found for the individual students, with a few exceptions. In particular, we observed that the presence of Socially Detached learners within a group did not significantly influence the group performance. This is most likely because, although they may be successful students individually, they do not engage meaningfully with their peers, and so have little impact on the group. These findings have implications for optimal group composition, suggesting that groups should not simply comprise high-participating members, but should include a combination of both low and high participators. However, what is perhaps even more important is that the group include members that are both aware of the social climate of the group interaction and invested in the collaborative outcome.

Another difference between the influence of roles on groups and individual performance pertains to the effect sizes. The influence of roles within a group appears to have a more potent influence on group performance (explaining 26%–28% of the variance) than does the influence of taking on a particular role on student performance (explaining only 2% of the variance). This illustrates the substantial impact that even a few members can have on a group, and the importance of diligent orchestration for optimal group composition.

Comparison to other group discourse modeling approaches

In addition to introducing and validating the GCA, we must also appropriately situate it within the literature with comparisons to other group discourse modeling approaches—namely,

to the contingency graph (Suthers, 2015; Suthers & Desiato, 2012), epistemic network analysis (ENA; Shaffer et al., 2009), and cohesion network analysis (CNA; Dascalu, McNamara, Trausan-Matu, & Allen, 2018). First, the contingency graph is used as the basis for representing transcriptions and highlights contingencies between events. The contingency graph relies on (a) events that can be traced to the interaction with the CSCL environment and (b) contingency relationships when one or more events enable a subsequent event. Some of the GCA measures share similarities with the contingency graph. Particularly, GCA places the same importance on the temporal and sequential nature of discourse, and the resulting GCA measures of responsivity, internal cohesion, social impact reflect in an automated manner the underlying contingency approach of capturing the micro-relations between situated discourse acts by participants, which are identified and then aggregated into interactional relations.

Second, Shaffer et al.’s (2009) ENA is rooted in a specific theory of learning: the epistemic frame theory, in which the collection of skill, knowledge, identity, value, and epistemology (SKIVE) forms an epistemic frame. A main theoretical assumption of ENA is that the connections between the elements of epistemic frames are critical for learning, not their presence in isolation. The online ENA toolkit allows users to analyze chat data by comparing the connections within the epistemic networks derived from chats. ENA visualization displays the clustering of learners and groups and the network connections of individual learners and groups. ENA requires coded data, which has traditionally relied on hand coded datasets or classifiers that rely on regular expression mapping. In contrast, GCA model is grounded in computational text analysis and semantic models that can facilitate a deeper understanding of discourse and the cohesive links among text segments. As such, GCA could be extended by utilizing the metrics and visualizations provided by ENA.

Dascalu et al.’s (2018) CNA forms an interesting alternative approach, which expands upon network analysis by explicitly considering semantic cohesion while modeling the interactions between participants. GCA shares some similar methodological inspiration with CNA, in that both GCA and CNA are grounded in semantic analysis that can facilitate understanding the cohesive links among text segments. However, the underlining theoretical and practical motivations of these approaches differ, and consequently the four metrics from CNA, and six metrics from GCA are not closely aligned. For example, CNA provides an Importance or Impact score, which in name appears similar to the GCA’s Social Impact measure. CNA’s cumulative importance score is derived from a mixture of both topic coverage and the existing cohesive links between contributions. By contrast, GCA is focused on capturing the intra- and inter-personal dynamics that reside in the discourse interactions between participants over time. Thus, GCA extracts sociocognitive process

measures, such as responsivity, internal cohesion, and social impact. CNA employs SNA methods to produce visually compelling sociograms. We feel that GCA could benefit from similar visualizations, especially those that illustrate the discourse dynamics and the resulting sociocognitive roles.

Conclusion and limitations

A primary objective of this research was to propose and validate a novel automated methodology, group communication analysis, for detecting emergent roles in group interactions. The GCA applies automated computational linguistic techniques to the sequential interactions of online collaborative interactions. The GCA involves computing six distinct measures of sociocognitive interaction patterns (i.e., Participation, Overall Responsivity, Social Impact, Internal Cohesion, Communication Density, and Sharing of New Information). The automated natural language metrics that make up the GCA provide a new and useful view on how roles are constructed and maintained in collaborative interactions.

There are some notable limitations to the variables selected for inclusion in the GCA. Particularly, the present research focused only on sociocognitive variables; however, several other collaborative interaction characteristics would likely provide valuable additional information as we attempt to characterize roles. For instance, the affective characteristics of individuals and groups have been shown to play a very important role in learning (Baker, D’Mello, Rodrigo, & Graesser, 2010; D’Mello & Graesser, 2012; Graesser, D’Mello, & Strain, 2014). There has also been evidence suggesting the importance of microbehavioral measures, such as keystrokes, click streams, response time, duration, and reading time measures, and these could provide additional information (Antonenko et al., 2012; Azevedo, Moos, Johnson, & Chauncey, 2010; Mostow & Beck, 2006). Finally, although we used the measure of topic relevance as an independent measure of group performance (i.e., separate from student learning gains) in the present work, this is arguably a feature that could provide valuable information for understanding social roles in group interactions. These limitations will be addressed in subsequent research.

One of the central contributions of the GCA can also be viewed as a limitation. One of the benefits of the preconceived categories involved in manual content analyses is that these coded categories afford a “gold standard” external validation. For instance, if these roles were identified through manually coded categories, then the cluster analysis results could be compared against the human-annotated “gold standard.” By pursuing a purely automated computational linguistic methodology, we were able to explore a substantially larger number of collaborative interactions than could be analyzed with

manual methods. Furthermore, given the complex and dynamic nature of the discourse characteristics that are calculated in the construction of the GCA, it would be extremely difficult and time consuming, if not impossible, for human coders to capture such multifaceted discourse characteristics. However, external cluster validation can be achieved either by comparing the cluster solutions to some “gold standard” categories or by comparing them to meaningful external variables (Antonenko et al., 2012). In the present research, we successfully took the latter approach, by showing that the identified roles are related to both individual student learning and group performance in general, and that the relationship is theoretically meaningful. Furthermore, even “gold standard” human coding schemes must be validated and tested for robustness. We feel that the tests of cluster stability, coherence, and internal consistency that we applied to our model are at least as extensive and rigorous as any interrater reliability study of a manual coding schema.

This research serves as an initial investigation with the GCA into understanding why some groups perform better than others. Despite some limitations, this research has provided some fruitful lines of research to be pursued in future work. Most significantly, the GCA provides us with a framework to investigate how roles are constructed and maintained through the dynamic sociocognitive processes within an interaction. Individual participants’ patterns of linguistic coordination and cohesion, as measured by the GCA, can diagnostically reveal the roles that individuals play in collaborative discussions. As a methodological contribution, therefore, we expect that the GCA will provide a more objective, domain-independent means for future exploration of roles than has been possible with manual coding rubrics. Moreover, as a practical contribution, substantially larger corpora of data can be analyzed with the GCA than when human time is required to annotate the data. Furthermore, the empirical findings of this research will contribute to our understanding of how individuals learn together as a group and thereby advance the cognitive, social, and learning sciences.

Author note This research was supported by the Army Research Institute (Grant W5J9CQ12C0043) and by the National Science Foundation (Grant IIS-1344257). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding agencies.

References

- Akkerman, S., Van den Bossche, P., Admiraal, W., Gijsselaers, W., Segers, M., Simons, R.-J., & Kirschner, P. (2007). Reconsidering group cognition: From conceptual confusion to a boundary area between

- cognitive and socio-cultural perspectives? *Educational Research Review*, 2, 39–63. doi:<https://doi.org/10.1016/j.edurev.2007.02.001>
- Alavi, M., & Dufner, D. (2004). Technology-mediated collaborative learning: A research perspective. In S. R. Hiltz & R. Goldman (Eds.), *Learning together online: Research on asynchronous learning networks* (pp. 191–213). Mahwah, NJ: Erlbaum.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46. doi:<https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, 60, 383–398.
- Arvaja, M., & Hämäläinen, R. (2008). Collaborative knowledge construction during structured tasks in an online course at higher education context. In G. F. Ollington (Ed.), *Teachers and teaching: Strategies, innovations and problem solving*. (pp. 359–376). New York, NY: Nova Science.
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45, 210–223. doi:<https://doi.org/10.1080/00461520.2010.515934>
- Azevedo, R., Winters, F. I., & Moos, D. C. (2004). Students collaboratively use hypermedia to learn science? The dynamics of self-and other-regulatory processes in an ecology classroom. *Journal of Educational Computing Research*, 31, 215–245. doi:<https://doi.org/10.2190/HFT6-8EB1-TN99-MJVQ>
- Bagley, E., & Shaffer, D. (2015). Learning in an urban and regional planning practicum: The view from educational ethnography. *Journal of Interactive Learning Research*, 26, 369–393.
- Baker, R. S. (2010). Data mining for education. In B. McGaw, P. Peterson, & E. Baker (Eds.), *International encyclopedia of education* (3rd ed., Vol. 7, pp. 112–118). Oxford, UK: Elsevier.
- Baker, R. S., D’Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human–Computer Studies*, 68, 223–241. doi:<https://doi.org/10.1016/j.ijhcs.2009.12.003>
- Bandura, A. (1994). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 61–96). Hillsdale, NJ: Erlbaum.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, 9, 403–436. doi:https://doi.org/10.1207/S15327809JLS0904_2
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12, 307–359. doi:https://doi.org/10.1207/S15327809JLS1203_1
- Beers, P. J., Boshuizen, H. P. A., Kirschner, P. A., & Gijsselaers, W. H. (2007). The analysis of negotiation of common ground in CSCL. *Learning and Instruction*, 17, 427–435. doi:<https://doi.org/10.1016/j.learninstruc.2007.04.002>
- Benamar, L., Balagué, C., & Ghassany, M. (2017). The identification and influence of social roles in a social media product community. *Journal of Computer-Mediated Communication*, 22, 337–362. doi:<https://doi.org/10.1111/jcc4.12195>
- Berkowitz, M. W., & Gibbs, J. C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, 399–410.
- Blei, D. M., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Blumenfeld, P. C., Kempler, T. M., & Krajcik, J. S. (2006). Motivation and cognitive engagement in learning environments. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 475–488). New York, NY, US: Cambridge University Press.
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R package for cluster validation. *Journal of Statistical Software*, 25, 4. doi:<https://doi.org/10.18637/jss.v025.i04>
- Cade, W. L., Dowell, N. M., Graesser, A. C., Tausczik, Y. R., & Pennebaker, J. W. (2014). Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 399–400). Berlin, Germany: Springer.
- Cai, Z., Burkett, C., Morgan, B., & Shaffer, D. (2011). Generation of domain-specific LSA spaces. Paper presented at the 41st Annual Meetings of the Society for Computers in Psychology (SCiP), Seattle, Washington.
- Cai, Z., Eagen, B., Dowell, N. M., Pennebaker, J. W., Shaffer, D., & Graesser, A. C. (2017). Epistemic network analysis and topic modeling for chat data from collaborative learning environment. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 104–111). Worcester, MA: International Educational Data Mining Society.
- Cai, Z., Li, H., Hu, X., & Graesser, A. C. (2016). Can word probabilities from LDA be simply added up to represent documents? In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 577–578). Worcester, MA: International Educational Data Mining Society. Retrieved from https://www.researchgate.net/publication/305045020_Can_word_probabilities_from_LDA_be_simply_added_up_to_represent_documents
- Çakır, M. P., Zemel, A., & Stahl, G. (2009). The joint organization of interaction within a multimodal CSCL medium. *International Journal of Computer-Supported Collaborative Learning*, 4, 115–149. doi:<https://doi.org/10.1007/s11412-009-9061-0>
- Capuano, N., Mangione, G. R., Mazzoni, E., Miranda, S., & Orciuoli, F. (2014). Wiring role taking in collaborative learning environments: SNA and semantic web can improve CSCL script? *International Journal of Emerging Technologies in Learning*, 9, 30–38.
- Care, Esther, Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, 29, 250–264. doi:<https://doi.org/10.1080/08957347.2016.1209204>
- Chan, C. K. K. (2012). Co-regulation of learning in computer-supported collaborative learning environments: a discussion. *Metacognition and Learning*, 7, 63–73. doi:<https://doi.org/10.1007/s11409-012-9086-z>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6), 1–36. doi:<https://doi.org/10.18637/jss.v061.i06>
- Chauhan, A. (2015). Beyond the phenomenon: Assessment in massive open online courses (MOOCs). In M. Elspeth (Ed.), *Macro-level learning through massive open online courses (MOOCs): Strategies and predictions for the future* (pp. 119–141). Hershey, PA: IGI Global. Retrieved from <https://doi.org/10.4018/978-1-4666-8324-2.ch007>
- Chi, M. T. H. (2009). Active–Constructive–Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105. doi:<https://doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns in peer collaboration that promote learning. In Lauren B. Resnick, C. S. C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 263–274). Washington, DC: American Educational Research Association. Retrieved from ebooks.aera.net/SOCCH21

- Chiu, M. M. (2000). Group problem-solving processes: Social interactions and individual actions. *Journal for the Theory of Social Behaviour*, 30, 26–49. doi:<https://doi.org/10.1111/1468-5914.00118>
- Clark, H. (1996). Using language. Cambridge, UK: Cambridge University Press.
- Clark, H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.
- Cress, U. (2008). The need for considering multilevel analysis in CSCL research—An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3, 69–84. doi:<https://doi.org/10.1007/s11412-007-9032-2>
- Dalton, L., Ballarin, V., & Brun, M. (2009). Clustering algorithms: On learning, validation, performance, and applications to genomics. *Current Genomics*, 10, 430–445. doi:<https://doi.org/10.2174/138920209789177601>
- Daradoumis, T., Martínez-Monés, A., & Xhafa, F. (2006). A layered framework for evaluating on-line collaborative learning interactions. *International Journal of Human-Computer Studies*, 64, 622–635. doi:<https://doi.org/10.1016/j.ijhcs.2006.02.001>
- Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L. K. (2018). Cohesion network analysis of CSCL participation. *Behavior Research Methods*, 50, 604–619. doi:<https://doi.org/10.3758/s13428-017-0888-4>
- Datta, S., & Datta, S. (2003). Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics*, 19, 459–466. doi:<https://doi.org/10.1093/bioinformatics/btg025>
- De Wever, B., Keer, H. V., Schellens, T., & Valcke, M. (2010). Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction. *Computers in Human Behavior*, 26, 516–523. doi:<https://doi.org/10.1016/j.chb.2009.08.008>
- De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2007). Applying multilevel modelling to content analysis data: Methodological issues in the study of role assignment in asynchronous discussion groups. *Learning and Instruction*, 17, 436–447. doi:<https://doi.org/10.1016/j.learninstruc.2007.04.001>
- Dennis, A. R. (1996). Information exchange and use in group decision making: You can lead a group to information, but you can't make it think. *MIS Quarterly*, 20, 433–457. doi:<https://doi.org/10.2307/249563>
- Dillenbourg, P. (1999). Collaborative learning: Cognitive and computational approaches. Amsterdam, The Netherlands: Pergamon.
- Dillenbourg, P., & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 21, 111–130.
- Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In D. N. Balacheff, D. S. Ludvigsen, D. T. de Jong, D. A. Lazonder, & D. S. Barnes (Eds.), *Technology-enhanced learning* (pp. 3–19). Amsterdam, The Netherlands: Springer. doi:https://doi.org/10.1007/978-1-4020-9827-7_1
- Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *Journal of the Learning Sciences*, 15, 121–151. doi:https://doi.org/10.1207/s15327809jls1501_9
- D'Mello, S., & Graesser, A. C. (2012). Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5, 304–317. doi:<https://doi.org/10.1109/TLT.2012.10>
- Doise, W. (1990). The development of individual competencies through social interaction. In H. C. Foot, M. J. Morgan, & R. Shute (Eds.), *Children helping children*. Chichester, UK: Wiley.
- Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies*, 26, 445–461. doi:<https://doi.org/10.1016/j.destud.2004.10.003>
- Dowell, N. M., Brooks, C., Kovanović, V., Joksimović, S., & Gašević, D. (2017). The changing patterns of MOOC discourse. In C. Urrea, J. Reich, & C. Thille (Eds.), *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale* (pp. 283–286). New York, NY: ACM.
- Dowell, N. M., Brooks, C., & Poquet, O. (2018). Applying group communication analysis to educational discourse interactions at scale. In J. Kay & R. Luckin (Eds.), *Proceedings of the 13th International Conference of the Learning Sciences* (pp. 1815–1822). London, UK: International Society of the Learning Sciences.
- Dowell, N. M., Cade, W. L., Tausczik, Y. R., Pennebaker, J. W., & Graesser, A. C. (2014). What works: Creating adaptive and intelligent systems for collaborative learning support. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Twelfth International Conference on Intelligent Tutoring Systems*. (pp. 124–133). Berlin, Germany: Springer.
- Dowell, N. M., & Graesser, A. C. (2015). Modeling learners' cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, 1, 183–186.
- Dowell, N. M., Skrypnik, O., Joksimović, S., Graesser, A. C., Dawson, S., Gašević, S., . . . Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In C. Romero & M. Pechenizkiy (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 250–257). Worcester, MA: International Educational Data Mining Society.
- Driskell, T., Driskell, J. E., Burke, C. S., & Salas, E. (2017). Team roles: A review and integration. *Small Group Research*, 48, 482–511. doi:<https://doi.org/10.1177/1046496417711529>
- Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4, 95–104. doi:<https://doi.org/10.1080/01969727408546059>
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Model-based cluster analysis for structured data. In *Cluster analysis* (pp. 187–213). Chichester, UK: Wiley. Retrieved from <http://onlinelibrary.wiley.com.proxy.lib.umich.edu/doi/10.1002/9780470977811.ch7/summary>
- Felps, W., Mitchell, T. R., & Byington, E. (2006). How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in Organizational Behavior*, 27, 175–222. doi:[https://doi.org/10.1016/S0191-3085\(06\)27005-9](https://doi.org/10.1016/S0191-3085(06)27005-9)
- Fiore, S., & Schooler, J. (2004). Process mapping and shared cognition: Teamwork and the development of shared problem models. In E. Salas & S. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 133–152). Washington, DC: American Psychological Association.
- Foltz, P. W., & Martin, M. J. (2009). Automated communication analysis of teams. In E. Salas, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pp. 411–431). Hove, UK: Psychology Press.
- Fox, J., & Weisberg, H. S. (2010). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal*, 35, 227–267. doi:<https://doi.org/10.2307/1163424>
- Gervits, F., Eberhard, K., & Scheutz, M. (2016). Team communication as a collaborative process. *Frontiers in Robotics and AI*, 3, 62. doi:<https://doi.org/10.3389/frobt.2016.00062>
- Gleave, E., Welsler, H. T., Lento, T. M., & Smith, M. A. (2009). A conceptual and operational definition of “social role” in online community. In *2009 42nd Hawaii International Conference on System*

- Sciences* (pp. 1–11). Piscataway, NJ: IEEE Press. doi:<https://doi.org/10.1109/HICSS.2009.6>
- Gorman, J. C., Cooke, N. J., & Kiekel, P. A. (2004). Dynamical perspectives on team cognition. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 673–677. doi:<https://doi.org/10.1177/154193120404800385>
- Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of latent semantic analysis-based measures of team communications content. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 47, 424–428. doi:<https://doi.org/10.1177/154193120304700337>
- Graesser, A. C., D’Mello, S. K., & Strain, A. C. (2014). Emotions in advanced learning technologies. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education*. New York, NY: Routledge. Retrieved from <https://www.routledgehandbooks.com/doi/10.4324/9780203148211.ch24>
- Graesser, A. C., Dowell, N., & Clewley, D. (2017). Assessing collaborative problem solving through conversational agents. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 65–80). New York, NY: Springer. doi:https://doi.org/10.1007/978-3-319-33261-1_5
- Graesser, A. C., Foltz, P., Rosen, Y., Shaffer, D., Forsyth, C., & Germany, M.-L. (2016). Challenges of assessing collaborative problem-solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills* (pp. 75–91). Heidelberg, Germany: Springer. doi:https://doi.org/10.1007/978-3-319-65368-6_5
- Gu, X., Shao, Y., Guo, X., & Lim, C. P. (2015). Designing a role structure to engage students in computer-supported collaborative learning. *The Internet and Higher Education*, 24, 13–20. doi:10.1016/j.iheduc.2014.09.002
- Guzdial, M., & Tums, J. (2000). Effective discussion through a computer-mediated anchored forum. *Journal of the Learning Sciences*, 9, 437–469.
- Gweon, G., Jain, M., McDonough, J., Raj, B., & Rosé, C. P. (2013). Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer-Supported Collaborative Learning*, 8, 245–265. doi:<https://doi.org/10.1007/s11412-013-9172-5>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64–74. doi:<https://doi.org/10.1119/1.18809>
- Han, J., Pei, J., & Kamber, M. (Eds.). (2012). *Data mining: Concepts and techniques*. Boston, MA: Elsevier.
- Hare, A. P. (1994). Types of roles in small groups: A bit of history and a current perspective. *Small Group Research*, 25, 433–448. doi:<https://doi.org/10.1177/1046496494253005>
- Hatano, G. (1993). Commentary: Time to merge Vygotskian and constructivist conceptions of knowledge acquisition. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children’s development* (pp. 153–166). New York, NY, US: Oxford University Press.
- Haviland, S. E., & Clark, H. H. (1974). What’s new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512–521. doi:[https://doi.org/10.1016/S0022-5371\(74\)80003-4](https://doi.org/10.1016/S0022-5371(74)80003-4)
- Hempelman, C. F., Duffy, D., McCarthy, P., Graesser, A. C., Cai, Z., & McNamara, D. S. (2005). Using LSA to automatically identify givenness and newness of noun-phrases in written discourse. In B. Bara (Ed.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 941–946). Mahwah, NJ: Erlbaum.
- Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, 52, 258–271. doi:<https://doi.org/10.1016/j.csda.2006.11.025>
- Hennig, C. (2015). fpc: Flexible procedures for clustering (Version 2.1-10). Retrieved from <https://cran.r-project.org/web/packages/fpc/index.html>
- Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of cluster analysis*. New York, NY: CRC Press.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Amsterdam, The Netherlands: Springer. doi:https://doi.org/10.1007/978-94-017-9395-7_2
- Hew, K. F., Cheung, W. S., & Ng, C. S. L. (2010). Student contribution in asynchronous online discussion: a review of the research and empirical exploration. *Instructional Science*, 38, 571–606. doi:<https://doi.org/10.1007/s11251-008-9087-0>
- Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, 26, 48–94. doi:<https://doi.org/10.1080/07370000701798495>
- Ho, S. M., Hancock, J. T., Booth, C., Burmester, M., Liu, X., & Timmarajus, S. S. (2016). Demystifying insider threat: Language-action cues in group dynamics. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 2729–2738). Piscataway, NJ: IEEE Press. doi:<https://doi.org/10.1109/HICSS.2016.343>
- Hoadley, C. (2010). Roles, design, and the nature of CSCL. *Computers in Human Behavior*, 26, 551–555. doi:<https://doi.org/10.1016/j.chb.2009.08.012>
- Hornik, K., & Böhm, W. (2016). clue: Cluster ensembles (Version 0.3-51). Retrieved from <https://cran.r-project.org/web/packages/clue/index.html>
- Howley, I., & Mayfield, E. (2011). Missing something? Authority in collaborative learning. In *Proceedings of the 9th International Computer Supported Collaborative Learning Conference* (Vol. 1, pp. 336–373). New York, NY: International Society of the Learning Sciences.
- Howley, I., Mayfield, E., & Rosé, C. P. (2013a). Linguistic analysis methods for studying small groups. In C. E. Hmelo-Silver, C. A. Chinn, & A. M. O’Donnell (Eds.), *The international handbook of collaborative learning* (pp. 184–203). New York, NY: Routledge.
- Howley, I., Mayfield, E., Rosé, C. P., & Strijbos, J.-W. (2013b). A multivocal process analysis of social positioning in study group interactions. In D. Suthers, K. Lund, C. P. Rose, C. Teplov, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions* (Vol. 15, pp. 205–223). New York, NY: Springer.
- Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, 51, 1755–1765. doi:<https://doi.org/10.1016/j.compedu.2008.05.005>
- Hu, X., Cai, Z., Wiemer-Hastings, P., Graesser, A. C., & McNamara, D. S. (2007). Strengths, limitations, and extensions of LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 401–426). New York, NY: Erlbaum.
- Hu, X., Zhang, X., Lu, C., Park, E. K., & Zhou, X. (2009). Exploiting Wikipedia as external knowledge for document clustering. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 389–396). New York, NY, USA: ACM. doi:<https://doi.org/10.1145/1557019.1557066>
- Hubert, L. J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218. doi:<https://doi.org/10.1007/BF01908075>
- Jahnke, I. (2010). Dynamics of social roles in a knowledge management community. *Computers in Human Behavior*, 26, 533–546. doi:<https://doi.org/10.1016/j.chb.2009.08.010>
- Janis, I. L. (1983). *Groupthink: Psychological studies of policy decisions and fiascoes*. Boston, MA: Houghton Mifflin.

- Janssen, J., Erkens, G., Kirschner, P. A., & Kanselaar, G. (2011). Multilevel analysis in CSCL research. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 187–205). New York, NY: Springer. doi:https://doi.org/10.1007/978-1-4419-7710-6_9
- Järvelä, S., Humme, T. R., & Järvelä, H. (2011). Self-regulation and motivation in computer supported collaborative learning environments. In S. Ludvigsen, S. Lund, I. Rasmussen, & R. Säljö (Eds.), *Learning across sites: New tools, infrastructures and practices* (pp. 330–345). New York, NY: Routledge.
- Joksimović, S., Dowell, N. M., Skypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). How do you connect? Analysis of social capital accumulation in connectivist MOOCs. In *Proceedings from the 5th International Learning Analytics and Knowledge (LAK) Conference* (pp. 66–68). New York, NY: ACM.
- Joksimović, S., Dowell, N. M., Poquet, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2018). Exploring development of social capital in a cMOOC through language and discourse. *Internet and Higher Education*, 36, 54–64. <https://doi.org/10.1016/j.iheduc.2017.09.004>
- Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation summarization in educational dialogue. In *Proceedings of the SLATE Workshop on Speech and Language Technology in Education* (pp. 53–56). Farmington, PA: ISCA.
- Kapur, M., Voiklis, J., & Kinzer, C. K. (2011). A complexity-grounded model for the emergence of convergence in CSCL groups. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 3–23). New York, NY: Springer. doi:https://doi.org/10.1007/978-1-4419-7710-6_1
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley-Interscience.
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive-load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31–42. doi:<https://doi.org/10.1007/s10648-008-9095-2>
- Kirschner, P. A., Beers, P. J., Boshuizen, H. P. A., & Gijsselaers, W. H. (2008). Coercing shared knowledge in collaborative learning environments. *Computers in Human Behavior*, 24, 403–420. doi:<https://doi.org/10.1016/j.chb.2007.01.028>
- Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. In W. B. Rouse & K. R. Boff (Eds.), *Organizational simulation* (pp. 139–184). Hoboken, NJ, USA: Wiley. doi:<https://doi.org/10.1002/0471739448.ch6>
- Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2015). MVN: Multivariate normality tests (Version 4.0). Retrieved from <https://cran.r-project.org/web/packages/MVN/index.html>
- Kotlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education*, 52, 132–142. doi:<https://doi.org/10.5032/jae.2011.01132>
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, 19, 335–353. doi:[https://doi.org/10.1016/S0747-5632\(02\)00057-2](https://doi.org/10.1016/S0747-5632(02)00057-2)
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer. Retrieved from <link.springer.com/10.1007/978-1-4614-6849-3>
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lehmann-Willenbrock, N., Beck, S. J., & Kauffeld, S. (2016). Emergent team roles in organizational meetings: Identifying communication patterns via cluster analysis. *Communication Studies*, 67, 37–57. doi:<https://doi.org/10.1080/10510974.2015.1074087>
- Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction*, 13, 487–509. doi:[https://doi.org/10.1016/S0959-4752\(02\)00042-7](https://doi.org/10.1016/S0959-4752(02)00042-7)
- Marcos-Garcia, J.-A., Martinez-Mones, A., & Dimitriadis, Y. (2015). DESPRO: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. *Computers & Education*, 82, 335–353. doi:<https://doi.org/10.1016/j.compedu.2014.10.027>
- McCarthy, P., Dufty, D., Hempelman, C. F., Cai, Z., Graesser, A. C., & McNamara, D. (2012). Newness and givenness of information: Automated identification in written discourse. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 475–478). Hershey, PA: IGI Global.
- Mesmer-Magnus, J. R., & Dechurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94, 535–546. doi:<https://doi.org/10.1037/a0013773>
- Mirriahi, N., Liaqat, D., Dawson, S., & Gašević, D. (2016). Uncovering student learning profiles with a video annotation tool: Reflective learning with and without instructional norms. *Educational Technology Research and Development*, 64, 1083–1106. doi:<https://doi.org/10.1007/s11423-016-9449-2>
- Molenaar, I., Chiu, M. M., Slegers, P., & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *International Journal of Computer-Supported Collaborative Learning*, 6, 601–624. doi:<https://doi.org/10.1007/s11412-011-9130-z>
- Mostow, J., & Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. *Natural Language Engineering*, 12, 195–208. doi:<https://doi.org/10.1017/S1351324906004153>
- Mudrack, P. E., & Farrell, G. M. (1995). An examination of functional role behavior and its consequences for individuals in group settings. *Small Group Research*, 26, 542–571. doi:<https://doi.org/10.1177/1046496495264005>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4, 133–142. doi:<https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Nolker, R. D., & Zhou, L. (2005). Social computing and weighting to identify member roles in online communities. In *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)* (pp. 87–93). Piscataway, NJ: IEEE Press. doi:<https://doi.org/10.1109/WI.2005.134>
- Nonnecke, B., & Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 73–80). New York, NY, USA: ACM. doi:<https://doi.org/10.1145/332040.332409>
- OECD. (2013). *PISA 2015 collaborative problem solving framework*. Oxford, UK: OECD.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2016). *vegan: Community ecology package* (Version 2.4-0). Retrieved from <https://cran.r-project.org/web/packages/vegan/index.html>
- Pennebaker, J. W., Gosling, S. D., & Ferrer, J. D. (2013). Daily online testing in large classes: Boosting college performance whereas reducing achievement gaps. *PLoS ONE*, 8, 79774. doi:<https://doi.org/10.1371/journal.pone.0079774>
- Piaget, J. (1993). *The moral judgement of the child*. New York, NY: Simon & Schuster.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects Models in S and S-Plus*. New York, NY: Springer.

- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., & Van Willigen, B. (2016). nlme: Linear and nonlinear mixed effects models (Version 3.1-128). Retrieved from <https://cran.r-project.org/web/packages/nlme/index.html>
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544–555. doi:<https://doi.org/10.1037/0022-0663.92.3.544>
- Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior, 20*, 201–223. doi:<https://doi.org/10.1016/j.chb.2003.10.015>
- Prince, E. F. (1981). Toward a taxonomy of given–new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York, NY: Academic Press.
- Risser, H. S., & Bottoms, S. (2014). “Newbies” and “celebrities”: Detecting social roles in an online network of teachers via participation patterns. *International Journal of Computer-Supported Collaborative Learning, 9*, 433–450. <https://doi.org/10.1007/s11412-014-9197-4>
- Roschelle, J. (1992). Learning by collaborating: convergent conceptual change. *Journal of the Learning Sciences, 2*, 235–276. doi:10.1207/s15327809jls0203_1
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem-solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 67–97). Berlin, Germany: Springer.
- Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*, 237–271. doi:<https://doi.org/10.1007/s11412-007-9034-0>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.
- Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine, 78*, B77–B85.
- Salazar, A. J. (1996). An analysis of the development and evolution of roles in the small group. *Small Group Research, 27*, 475–503. doi:<https://doi.org/10.1177/1046496496274001>
- Samsonovich, A. V. (2014). Semantic cross-correlation as a measure of social interaction. *Biologically Inspired Cognitive Architectures, 7*(Suppl. C), 1–8. doi:<https://doi.org/10.1016/j.bica.2013.12.001>
- Sapru, A., & Bourlard, H. (2015). Automatic recognition of emergent social roles in small group interactions. *IEEE Transactions on Multimedia, 17*, 746–760. doi:<https://doi.org/10.1109/TMM.2015.2408437>
- Sarmiento, J. W., & Shumar, W. (2010). Boundaries and roles: Positioning and social location in the Virtual Math Teams (VMT) online community. *Computers in Human Behavior, 26*, 524–532. doi:<https://doi.org/10.1016/j.chb.2009.08.009>
- Shaffer, D. (2006). *How computer games help children learn*. New York, NY: Palgrave Macmillan.
- Shaffer, D., & Graesser, A. C. (2010). Using a quantitative model of participation in a community of practice to direct automated mentoring in an ill-defined domain. In C. Lynch, K. Ashley, T. Mitrovic, V. Dimitrova, N. Pinkwart, & V. Aleven (Eds.), *Proceedings of the 4th International Workshop on Intelligent Tutoring Systems and Ill-Defined Domains* (pp. 61–68). Pittsburgh, PA: Workshop on Intelligent Tutoring Systems (ITS).
- Shaffer, D., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., . . . Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media, 1*, 33–53. doi:<https://doi.org/10.1162/ijlm.2009.0013>
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 1. Retrieved from http://itdl.org/Journal/Jan_05/article01.htm
- Spada, H. (2010). Of scripts, roles, positions, and models. *Computers in Human Behavior, 26*, 547–550. doi:<https://doi.org/10.1016/j.chb.2009.08.011>
- Stahl, G. (2002). Rediscovering CSCL. In T. Koschmann, R. Hall, & N. Miyake (Eds.), *CSCL 2: Carrying forward the conversation* (pp. 169–181). Mahwah, NJ: Erlbaum.
- Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning, 21*, 79–90. doi:<https://doi.org/10.1111/j.1365-2729.2005.00115.x>
- Stahl, G. (2010). Group cognition as a foundation for the new science of learning. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning* (pp. 23–44). New York, NY: Springer. doi:https://doi.org/10.1007/978-1-4419-5716-0_2
- Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning, 8*, 1–12. doi:<https://doi.org/10.1007/s11412-013-9169-0>
- Stahl, G., Law, N., Cress, U., & Ludvigsen, S. (2014). Analyzing roles of individuals in small-group collaboration processes. *International Journal of Computer-Supported Collaborative Learning, 9*, 365–370. doi:<https://doi.org/10.1007/s11412-014-9204-9>
- Stahl, G., & Rosé, C. P. (2013). Theories of team cognition: Cross-disciplinary perspectives. In E. Salas, S. M. Fiore, & M. P. Letsky (Eds.), *Theories of team cognition: Cross-disciplinary perspectives* (pp. 111–134). New York, NY: Routledge.
- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*, 1467–1478. doi:<https://doi.org/10.1037/0022-3514.48.6.1467>
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning* (pp. 1–15). Mahwah, NJ: Erlbaum.
- Strijbos, J.-W., & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An explorative synthesis. *Computers in Human Behavior, 26*, 495–505. doi:<https://doi.org/10.1016/j.chb.2009.08.014>
- Strijbos, J.-W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The effect of functional roles on group efficiency using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research, 35*, 195–229. doi:<https://doi.org/10.1177/1046496403260843>
- Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29–48. doi:<https://doi.org/10.1016/j.compedu.2005.04.002>
- Strijbos, J.-W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*, 491–494. doi:<https://doi.org/10.1016/j.chb.2009.08.006>
- Stuetzer, C. M., Koehler, T., Carley, K. M., & Thiem, G. (2013). “Brokering” behavior in collaborative learning systems. *Procedia: Social and Behavioral Sciences, 100*, 94–107. doi:<https://doi.org/10.1016/j.sbspro.2013.10.702>
- Suthers, D. (2006). A qualitative analysis of collaborative knowledge construction through shared representations. *Research and Practice in Technology Enhanced Learning, 1*, 315–337. doi:<https://doi.org/10.1142/S1793206806000147>
- Suthers, D. (2015). From contingencies to network-level phenomena: Multilevel analysis of activity and actors in heterogeneous networked learning environments. In *Proceedings of the Fifth*

- International Conference on Learning Analytics and Knowledge* (pp. 368–377). New York, NY, USA: ACM. doi:<https://doi.org/10.1145/2723576.2723626>
- Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *2012 45th Hawaii International Conference on System Sciences* (pp. 3368–3377). Piscataway, NJ: IEEE Press. doi:<https://doi.org/10.1109/HICSS.2012.274>
- Suthers, D., Dwyer, N., Medina, R., & Vatrappu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, 5, 5–42. doi:<https://doi.org/10.1007/s11412-009-9081-9>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson.
- Taniar, T. (2006). *Research and trends in data mining technologies and applications*. London, UK: Idea Group.
- Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In Lauren B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning* (pp. 361–384). Berlin, Germany: Springer. doi:https://doi.org/10.1007/978-3-662-03362-3_16
- Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, 14, 511–528. doi:<https://doi.org/10.1198/106186005X59243>
- Valle, R. del, & Duffy, T. M. (2007). Online learning: Learner characteristics and their approaches to managing learning. *Instructional Science*, 37, 129–149. doi:<https://doi.org/10.1007/s11251-007-9039-0>
- Van Boxtel, C. (2004). Studying peer interaction from three perspectives. In J. van der Linden & P. Renshaw (Eds.), *Dialogic learning* (pp. 125–143). New York, NY: Springer. doi:https://doi.org/10.1007/1-4020-1931-9_7
- Van den Bossche, P., Gijsselaers, W. H., Segers, M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments team learning beliefs and behaviors. *Small Group Research*, 37, 490–521. doi:<https://doi.org/10.1177/1046496406292938>
- Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, 19, 128–143. doi:<https://doi.org/10.1016/j.learninstruc.2008.03.001>
- Volet, S., Vauras, M., Salo, A.-E., & Khosa, D. (2017). Individual contributions in student-led collaborative learning: Insights from two analytical approaches to explain the quality of group outcome. *Learning and Individual Differences*, 53, 79–92. doi:<https://doi.org/10.1016/j.lindif.2016.11.006>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
- Wen, M., Yang, D., & Rosé, C. (2014). Linguistic reflections of student engagement in massive open online courses. In *Proceedings of the 14th International Conference on Web and Social Media* (pp. 525–534). Ann Arbor, MI: AAAI. Retrieved from <http://meefen.github.io/notes/2014/05/01/Wen2014/>
- Winne, P. H. (2013). Metacognition in computer-supported collaborative inquiry. In C. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 462–479). New York, NY: Taylor & Francis.
- Wise, A. F., Speer, J., Marbouti, F., & Hsiao, Y.-T. (2012). Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors. *Instructional Science*, 41, 323–343. doi:<https://doi.org/10.1007/s11251-012-9230-9>
- Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: moving beyond the hidden profile paradigm. *Communication Monographs*, 71, 286–310. doi:<https://doi.org/10.1080/0363452042000299894>
- Wittenbaum, G. M., & Stasser, G. (1996). Management of information in small groups. In J. L. Nye & A. M. Brower (Eds.), *What's social about social cognition? Research on socially shared cognition in small groups* (pp. 3–28). Thousand Oaks, CA: Sage. Retrieved from <hdl.handle.net/2027/>
- Xie, K., Yu, C., & Bradshaw, A. C. (2014). Impacts of role assignment and participation in asynchronous discussions in college-level online classes. *The Internet and Higher Education*, 20, 10–19. doi:<https://doi.org/10.1016/j.iheduc.2013.09.003>
- Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum. Retrieved from <http://hdl.handle.net/2027/>
- Zumel, N., Mount, J., & Porzak, J. (2014). *Practical data science with R*. Shelter Island, NY: Manning.