

# Exploring the Relationship Between Emergent Sociocognitive Roles, Collaborative Problem-Solving Skills, and Outcomes: A Group Communication Analysis

Nia M. M. Dowell<sup>1</sup>, Yiwen Lin<sup>2</sup>, Andrew Godfrey<sup>3</sup>, and Christopher Brooks<sup>4</sup>

## Abstract

Collaborative problem-solving (CPS) has become an essential component of today's knowledge-based, innovation-centred economy and society. As such, communication and CPS are now considered critical 21st century skills and incorporated into educational practice, policy, and research. Despite general agreement that these are important skills, there is less agreement on how to capture sociocognitive processes automatically during team interactions to gain a better understanding of their relationship with CPS outcomes. The availability of naturally occurring educational discourse data within online CPS platforms presents a golden opportunity to advance understanding about online learner sociocognitive roles and ecologies. In this paper, we explore the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes. Group Communication Analysis (GCA) — a computational linguistic framework for analyzing the sequential interactions of online team communication — was applied to a large CPS dataset in the domain of science (participant N = 967; team N = 480). The ETS Collaborative Science Assessment Prototype (ECSAP) was used to measure learners' CPS skills, and CPS outcomes. Cluster analyses and linear mixed-effects modelling were used to detect learner roles, and assess the relationship between those roles on CPS skills and outcomes. Implications for future research and practice are discussed regarding sociocognitive roles and collaborative problem-solving skills.

## Notes for Practice

- Through machine learning models, GCA overcomes some of the constraints inherent to manual qualitative analysis or summative data streams and provides novel perspectives by quantifying intra- and interpersonal dynamics involved in CPS.
- GCA provides educational stakeholders with a domain-independent scalable framework to investigate how roles are constructed and maintained through the dynamic sociocognitive processes within STEM CPS interactions. Further, the novel sociocognitive dimensions within GCA provide new indicators of CPS outcomes and learner CPS competencies not captured in current practices.
- The classification of types of CPS skills and behaviours into a smaller set of sociocognitive roles reduces the analytic complexity of capturing CPS skills within a social system and facilitates the comparative study of populations across time and setting

## Keywords

Group communication analysis, sociocognitive roles, computational linguistics, collaborative problem-solving, STEM

**Submitted:** 08/05/2019 — **Accepted:** 28.01.2020 — **Published:** 03.04.2020

Corresponding author<sup>1</sup> Email: [nia.dowell@uci.edu](mailto:nia.dowell@uci.edu) Address: School of Education, University of California, Irvine, 3200 Education Building, Irvine, California, 92697, ORCID ID: <https://orcid.org/0000-0002-9839-8947>

<sup>2</sup> Email: [yiwenl21@uci.edu](mailto:yiwenl21@uci.edu) Address: School of Education, University of California, Irvine, 3200 Education Building, Irvine, California, 92697. ORCID ID: <https://orcid.org/0000-0003-3602-8454>

<sup>3</sup> Email: [andgodfr@umich.edu](mailto:andgodfr@umich.edu) Address: Department of Psychology, University of Michigan, 500 S. State St. #2005, Ann Arbor, MI 48109. ORCID ID: <https://orcid.org/0000-0002-7807-7011>

<sup>4</sup> Email: [brookschr@umich.edu](mailto:brookschr@umich.edu) Address: School of Information, University of Michigan, 500 S. State St. #2005, Ann Arbor, MI 48109, ORCID ID: <https://orcid.org/0000-0003-0875-0204>

## 1. Introduction

Virtual teams are increasingly being relied upon for collaborative problem-solving across domains (Fiore, Graesser, & Greiff, 2018; Graesser et al., 2018b). To contend with dynamic conditions, communication and collaborative problem-solving (CPS) skills have become essential components of today's economy and society. As such, improving these competencies has taken a principal role in educational policy, research, and technology (Binkley et al., 2011; Care, Scoular, & Griffin, 2016; Graesser et al., 2018c; Graesser, Dowell, & Clewley, 2017; National Research Council, 2011; OECD, 2013; von Davier, Hao, Liu, & Kyllonen, 2017).

The importance of CPS skills is reflected in the evolution of higher education where the focus of learning and problem-solving has been evolving from traditional, highly individualistic processes and products to a more interactive process involving teams and social networks (Binkley et al., 2011; Dede, 2010; Freeman et al., 2014; Voogt, Erstad, Dede, & Mishra, 2013). However, current assessment practices of CPS processes and products have been increasingly criticized (von Davier et al., 2017). Researchers traditionally rely on questionnaires or individual perceptions to evaluate cognitive and social aspects of CPS. While these approaches can provide valuable information, they often lack generalizability and afford only a limited understanding around the actual process of collaboration. In order to move forward, the field is in need of interdisciplinary efforts focused on creating domain-independent methods that allow us to identify and characterize the complex mechanisms involved in collaboration.

Over the past decade, technology-enriched CPS environments have become pervasive in both formal and informal educational contexts (Greenhow, Robelia, & Hughes, 2009). This is evident in emerging educational trends towards computer-mediated CPS environments (Dillenbourg, Järvelä, & Fischer, 2009; Jeong & Hmelo-Silver, 2016), single conversational computer agents in CPS (Graesser et al., 2018a), and multi-agent and human CPS interactions (Hao, Liu, von Davier, & Kyllonen, 2015, 2017; Hu et al., 2018). These educational CPS technologies, when leveraged appropriately, provide several opportunities for the learning analytics community to make critical strides in advancing the science of STEM CPS (Graesser et al., 2018a). The insulated nature of digital CPS environments allows valuable problem-solving dynamics to be captured through learners' trace and discourse data at a high resolution and scale. These dynamics provide a window into learners' social, cognitive, and affective processes and their relationship to CPS outcomes. Furthermore, learners' trace and discourse data fit the criteria in affording a means for robust and domain-independent CPS assessments. Such learning analytics approaches to CPS investigations have proven successful in detecting and predicting the quality of team-based work (Guo, Yu, Pearlman, Navab, & Barmaki, 2019).

In this work, we are interested in characterizing team communication dynamics to identify the roles that learners occupy during technology-mediated STEM CPS interactions and understand the relationship between those roles and CPS skills and outcomes. Towards this effort, we employ Group Communication Analysis (GCA, described in more detail below; Dowell, Nixon, & Graesser, 2018), a computational linguistics methodology for quantifying the discourse dynamics between learners in online multi-party interactions. GCA calculates six inter- and intrapersonal sociocognitive measures of user interactions, and from these identifies distinct interaction profiles through a cluster analysis. Using this approach, we are able to gain a deeper understanding of learners' emergent roles in relation to their CPS skills and outcomes to clarify their corresponding relationships.

The subsequent sections of the paper are organized as follows. First, we describe the conceptual foundations of CPS assessment and the GCA measures situated within the context of CPS literature. We then provide a discussion around GCA and the framework for the assessment of CPS skills adopted in this study. Second, we discuss the importance of sociocognitive roles with CPS, before moving into the methodological features of the current investigation, followed by the details of the cluster analysis used to identify specific learner profiles in the communication patterns during CPS interactions. Third, we discuss the linear mixed-effects modelling used to explore the relationship between the identified sociocognitive roles/learners and group CPS performance/CPS skills. Finally, we conclude the paper 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 problem-solving interaction research.

### 1.1. Collaborative Problem-Solving: A Sociocognitive View

Collaboration and CPS have garnered significant research attention, with efforts directed towards understanding the processes involved in, and measurement of CPS skills and outcomes (von Davier et al., 2017). This has produced several empirical studies, and theoretical models outlining the individual behaviours and team mechanisms involved in effective collaborative problem-solving (Care et al., 2016; Dillenbourg & Fischer, 2007; Dillenbourg & Traum, 2006; Graesser et al., 2018a, 2018c; Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Howley & Rosé, 2016; Kirschner, Beers, Boshuizen, & Gijssels, 2008; Malmberg, Järvelä, & Järvenoja, 2017; OECD, 2013; Roschelle & Teasley, 1995; Rosé et al., 2014; Siddiq & Scherer, 2017; Teasley et al., 2008; Teasley & Roschelle, 1993; von Davier & Halpin, 2013).

Within this literature, there is a consistent theme highlighting the significance of both social and cognitive communication factors in facilitating problem-solving and learning in teams. Although popularized work such as the PISA framework offers effective constructs for evaluating CPS practices, there is a tendency to view the social and cognitive realm as either a dichotomy or one in the same (i.e., sociocultural perspective). We suggest that a more appropriate conception of these constructs is a two-dimensional space, highlighting the tightly coupled nature of cognition and social processes. Indeed, several writers contend that many CPS practices — such as sharing and negotiating ideas and prior experiences as well as jointly regulating and coordinating behaviours — are sustained through and fundamentally rely on social strategies (Dillenbourg et al., 2009; Hao, Liu, von Davier, & Kyllonen, 2017; Järvelä, Volet, & Järvenoja, 2010; van den Bossche, Segers, & Kirschner, 2006; von Davier et al., 2017). We have observed that the operationalization of learners' cognitive engagement in CPS has also extended from an individual's ability to plan, set goals, and complete assigned tasks to their reflective and regulatory processes involved in social interaction, further suggesting the intertwined nature of social and cognitive constructions in collaborative environments (Järvelä et al., 2010). From this view, it would seem that CPS practices could best be described as sociocognitive in nature, lying within a social/cognitive two-dimensional space, as opposed to being one in the same or dichotomous since CPS outcomes generally result from the interaction of both (Hao, Liu, von Davier, & Kyllonen, 2017). Capturing the sociocognitive communication patterns of language between interacting individuals during CPS is of great value. However, until recently quantifying sociocognitive properties through meaningful natural language patterns of learners has proven challenging (Care et al., 2016).

Group communication analysis (GCA) was motivated by the need for automated approaches to characterize sociocognitive discourse mechanisms within computer-mediated communication (Dowell, 2017; Dowell, Nixon, & Graesser, 2018; Dowell, Poquet, & Brooks, 2018; Dowell, 2019; Dowell, Lin, Godfrey, Cho, & Brooks, 2019; Lin, Dowell, Godfrey, Cho, & Brooks, 2019). The sociocognitive aspects of human interaction reside in and evolve through the semantic connection between individuals' utterances over time. GCA employs automated computational linguistic analysis to the sequential interactions of participants utterances during online conversations. This approach affords a mapping of the relational structure or semantic connection between learners' contributions over time in online group communication. As such, GCA both captures the structure of the group discussion, and quantifies the complex semantic cohesion relationships between learners' contributions over time, revealing intra- and interpersonal processes in group communication. Specifically, GCA calculates six inter- and intrapersonal sociocognitive measures of multiparty interactions (i.e., 1) participation, 2) overall responsivity, 3) internal cohesion, 4) social impact, 5) newness, and 6) communication density). Table 1 provides a description of these six GCA dimensions, and more detail can be found in Dowell (2017) and Dowell, Nixon, & Graesser (2018).

GCA is grounded upon theoretical constructs from several areas where sociocognitive processes, team interaction, and collaborative skills have been assessed. As GCA approaches interactions in a domain-independent manner, it offers implications on a variety of collaborative activities (Dowell, 2017; Dowell, Nixon, & Graesser, 2018). In the context of CPS, modelling group dynamics with GCA provides opportunities to examine how aspects of team interaction and individual roles influence learners' collaboration strategies, such as negotiation and regulation, and CPS outcomes.

## 1.2. Conceptual Alignment: GCA and CPS skills

In this study, we adopted the CPS framework proposed by Liu, Hao, von Davier, Kyllonen, and Zapata-Rivera (2016), which was developed based on the findings from computer-supported collaborative learning (CSCL) research (Barron, 2003; Dillenbourg & Traum, 2006) and the PISA 2015 Collaborative Problem Solving Framework (OECD, 2013). In this CPS framework, the authors present problem-solving competence as a spectrum in-between social and cognitive factors. Four main CPS skills are targeted by the framework: 1) sharing ideas, 2) negotiating ideas, 3) regulating problem-solving activities, and 4) maintaining communication (Table 1). Each of these major categories has subcategories, yielding a total of 33 subcategories. (For a more detailed discussion of the rubric, see Hao, Liu, von Davier, & Kyllonen, 2017; Liu et al., 2016.) Figure 1 illustrates GCA dimensions metrics in position relative to the four categories of CPS skills.

In terms of GCA measures, participation has a beneficial influence on perceived and actual learning, retention rates, learner satisfaction, social capital, and reflection (Hew, Cheung, & Ng, 2010; for a review see Hrastinski, 2008). Posting a message to the CPS environment is often operationalized by researchers and instructors as participation (Hrastinski, 2008) and considered a requirement for sharing and maintaining communication in any online learning group. Within CPS, participation is regarded as a signifier of willingness and readiness of learners to externalize and share information with others, but does not account for the semantic relatedness between utterances or behaviours (Hesse et al., 2015). As such, participation is situated on the more social end of the sociocognitive spectrum in Figure 1.

GCA's social impact and overall responsivity dimensions capture the content relatedness in learner participation and, as such, are inherently more cognitive than mere participation. Social impact quantifies the extent to which a learner's utterances are perceived as meaningful, or worthy of further discussion, by their teammates by measuring the semantic relatedness

between the participant’s current contribution and those that follow from their collaborative partners. For instance, learners may acquire higher degrees of social impact when they make contributions that are particularly insightful for the problem-solving task. Higher social impact could correspondingly be associated with more dysfunctional behaviour, as in the case of antagonistic team members. In both of these instances (i.e., productive or detrimental), social impact involves some degree of social monitoring for higher degrees to be achieved.

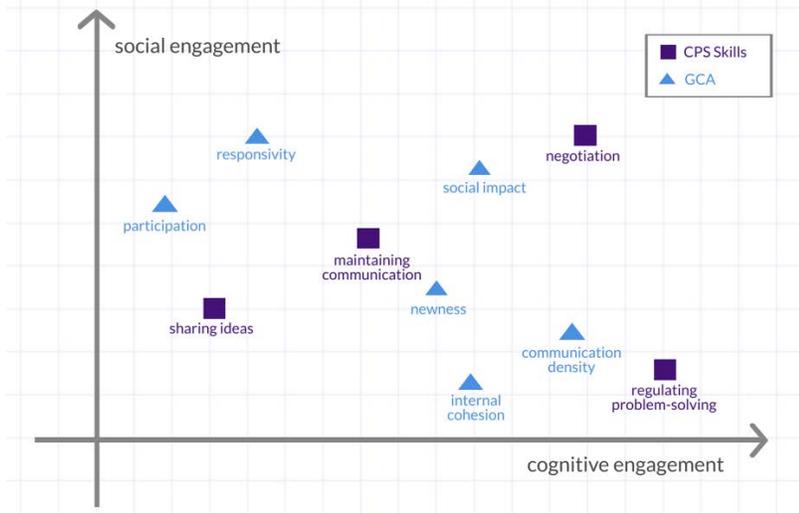


Figure 1. Visualization of GCA measures relative to CPS skills within a two-dimensional sociocognitive space. Note that position along the spectrum was not empirically obtained and is intended to be illustrative and not authoritative.

Table 1. GCA Measures and CPS Skills

Measure	Description
<b>GCA Measures</b>	
Participation	Mean participation of any participant above or below what you would expect from equal participation in a group of their size
Social Impact	Measure of how contributions initiated by the corresponding participant have triggered follow-up responses
Overall Responsivity	Measure of how responsive a participant’s contributions are to all other group members’ recent contributions
Newness	The amount of new information a participant provides, on average
Internal Cohesion	How semantically similar a participant’s contributions are with their own recent contributions
Communication Density	The amount of semantically meaningful information
<b>CPS Skills</b>	
Sharing Ideas	Presenting or responding to a need for task-relevant information
Negotiating Ideas	Discussing differences and similarities in presented ideas and addressing agreement and disagreement
Regulating Problem-Solving Activities	Leading discussion by evaluating current group interactions and progress
Maintaining Communication	Social oriented communication that may or may not be task relevant

Overall responsivity captures the semantic relatedness between the individual’s current contribution and the previous contributions of their teammates (i.e., uptake; Suthers et al., 2010). This information can be reflective of what has been referred to as knowledge convergence, or common ground (Clark, 1996; Clark & Brennan, 1991; Fiore & Schooler, 2004; Roschelle, 1992; Teasley & Roschelle, 1993). In the CPS literature, it is recognized that learners must also monitor and build on the perspectives of their teammates to achieve and maintain a shared understanding of the task and its solutions (Dillenbourg & Traum, 2006; Graesser et al., 2018c; Hmelo-Silver & Barrows, 2008; OECD, 2013; Stahl & Rosé, 2013). Taken together, we

see that social impact and overall responsivity are relational constructs that signify a connection between an individual's internal cognitive process and communicative social elements external to them.

The distinction between given (old) information versus new information in discourse is a foundational distinction in theories of discourse processing (Price, 1981). **Newness** measures the amount of new information in a conversation, and involves words, concepts, and ideas not yet mentioned, and builds on the given information or begins a new thread of ideas. In the context of CPS, peer interactions 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. However, despite the intuitive importance of (new) information sharing, a consistent finding from research is that groups predominantly discuss information that has been already shared (known to all participants) at the expense of information that has not been shared (known to a single member; see Mesmer-Magnus & Dechurch, 2009, for a review).

While collaboration is understood as an unequivocally interpersonal and contextual phenomenon, we argue that the role of an individual interacting with themselves should also be taken into account (Barron, 2000; Kuhn, 2015; OECD, 2013; Stahl, 2002). The process of individuals engaging in self-monitoring and reflection is captured by **internal cohesion**. During peer problem-solving interactions, this is necessary for individuals to appropriately build on and integrate their own views with those of the group (Kreijns, Kirschner, & Jochems, 2003; OECD, 2013). **Communication density** measures the ratio of semantic content to the number of words used to convey that content. It is suggested that concise communication indicates more common ground within the team and the presence of shared mental models of the task and team interaction (Klein, Feltovich, Bradshaw, & Woods, 2005). Communicating through clear language is also conducive to spearheading towards solutions for the whole team.

In summary, both the GCA and CPS frameworks capture the interwoven processes of social and cognitive engagement. Taken together, successful CPS is achieved through effective 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 towards the solution. These activities are supported in several collaborative learning perspectives (e.g., cognitive elaboration [Chi, 2009]; socio-cognitive conflict [Doise, 1990]; co-construction [Hatano, 1993; Piaget, 1993; Van Boxtel, 2004]), each of which stress different mechanisms to facilitate progress during group interactions (giving, receiving, and using explanations, resolving conflicts, co-construction). However, little is known about how these composite sociocognitive processes interact, progress, and influence CPS skills and outcomes. In the following section, we discuss how GCA can unpack these intricate relationships through the lens of roles and provide implications for CPS performance.

### 1.3. The Role: Quantifying Patterns of Behaviour

Research in CPS has historically focused on assessing important aspects such as individual behaviours, communication, and collaboration competencies. These frames prioritize research detailing the characteristics and behaviours of interacting individuals, while the social and cognitive identity concepts that lead to intra- and interindividual and group dynamics have received far less attention. There has been growing interest in understanding how individuals personalize their efforts during CPS (Singley, Fairweather, & Swerling, 1999) in online communities and learning interactions more broadly (Dowell, Nixon, & Graesser, 2018; Dowell, Poquet, & Brooks, 2018; Gleave, Welsler, Lento, & Smith, 2009; Yeh, 2010). Essentially, this issue involves distinguishing between a focus on competence, or the extent of one's communication and individual CPS skills in isolation, and a focus on mode or preference for certain ways of expressing or engaging in CPS interactions across a range of behavioural and relational dimensions.

There are two main perspectives on roles in the literature, namely scripted or assigned roles, and emergent roles (Cesareni, Cacciamani, & Fujita, 2015). While there have been research efforts directed towards investigating aspects of scripted roles (e.g., Gervits, Eberhard, & Scheutz, 2016), emergent roles in CPS have received far less attention. In the current research, we are most interested in the emergent roles that learners occupy during CPS interactions. In contrast to scripted roles, emergent roles are conceptualized as a product of a specific interactional context, consisting of patterns in sociocognitive behaviours enacted by people (Dowell, Nixon, & Graesser, 2018; Gleave et al., 2009; Strijbos & Weinberger, 2010). Accordingly, a role is an upper-level concept of a behaviour and can comprise a set of behaviours. Roles are emergent in that they develop naturally out of the interpersonal interaction over time without any prior instruction or assignment, and are defined (characterized) by their behavioural proximity (similarities and differences) to other interactional partners.

The roles that learners occupy during CPS can influence both CPS skills and outcomes (Dornfeld & Puntambekar, 2015; Spada, 2010). Successful CPS interactions can be threatened by a social loafer, or uncooperative participant, whereas it can be facilitated by a strong team member who pulls from different perspectives, helps negotiate conflicts, promotes team communication, and guides the group to overcome difficult obstacles (Fiore et al., 2010; Graesser et al., 2018b; Letsky, 2008; Salas, Cooke, & Rosen, 2008). Despite the importance of emergent roles in CPS contexts, there have been limited efforts

towards characterizing learners' emergent roles. In the current research, we use the GCA (described in section 1.2) to identify the emergent roles learners occupy during CPS. Using this approach, we are able to gain a deeper understanding of emergent roles in relation to CPS skills and outcomes to clarify their corresponding relationships.

#### 1.4. Overview of the Present Research

Drawing on this, we explore learner discourse during collaborative problem-solving interactions by employing the Group Communication Analysis (GCA). As the GCA both captures the structure of team discussions, and quantifies the complex semantic cohesion relationships between learner contributions over time, it can be used to characterize the **social roles** individuals occupy in CPS team interactions. In doing so, this methodology goes beyond previous models for automated group communication, which often rely on quantifying the structural components of human interaction without taking into account the semantic content of the discourse being exchanged between learners (e.g. social network analysis).

The current research has three main objectives. The first is to detect emergent roles in learning-based collaborative problem-solving interactions using the GCA. The second is to explore how the individual-level roles and overall team compositions influence both student and team performance during collaborative problem-solving interactions. The final objective is to investigate the association between the identified sociocognitive roles and learners' CPS skills. The concepts, methods, and ideas presented in this research are at the intersection of collaborative interaction, discourse processes, data mining, and learning analytics. This interdisciplinary research approach will hopefully provide insights regarding the nature of social identity in team interactions.

#### 1.5 Research Questions

RQ1. What sociocognitive roles do learners take on during collaborative problem-solving interactions?

RQ2. How do emergent roles influence individual and team problem-solving performance?

RQ3. What is the relationship between an individual's role and their collaborative problem-solving skills and proficiencies?

## 2. Methods

### 2.1. Participants

Participants in the current study included  $N = 967$  individuals collected in a previous study (Hao, Liu, von Davier, & Kyllonen, 2017; Liu et al., 2016). Students were recruited using Amazon Mechanical Turk, a crowdsourcing data collection platform (Kittur, Chi, & Suh, 2008). All participants had at least one year of college experience. Participants were randomly paired into dyads to take part in a tetralogue-based CPS task. After removing incomplete responses, we had data from 480 dyads. An approximately equal number of males and females participated, and their age ranged from 18 to 68 ( $M = 30.12$ ). Of those participants who reported their race and ethnicity (877 out of 967), 78% of the participants were white, 7% were black or African American, 5% were Asian, 5% were Hispanic or Latino, and 5% were multiracial.

### 2.2. Procedure

The ETS Platform for Collaborative Assessment and Learning (EPCAL) was used as a platform for this study. The ETS Collaborative Science Assessment Prototype (ECSAP) was used to measure learners' CPS skills, personality, and CPS outcomes (Hao et al., 2017). Prior to completing the CPS task, participants individually completed a science content test to assess their science knowledge, a background questionnaire, and a personality questionnaire. The background questionnaire asked for participants' race, gender, age, highest level of education completed, native language, computer use, and parents' highest level of education, for example.

Participants completed a simulation-based collaborative problem-solving task in which two individuals worked together in a virtual laboratory to solve a science problem about making predictions about volcanic eruptions. The simulation task was modified from an existing simulation, Volcano Dialogue (Zapata-Rivera et al., 2014). Two participants were automatically formed into a dyad based on whether they logged into the system within a similar time range. The dyads worked together, for an average of 60 minutes, via text chat to complete the tasks. A "turn" consists of whatever text a participant types before pressing "send." About 80% of the sessions had 35–100 turns. The chattiest session had 300 turns. Overall, there are 38,703 turns in our corpus. The total number of tokens is 189K (213K with punctuation). Average token-count per turn is 4.9 (5.5 with punctuation).

Participants were guided by a virtual scientist agent who presented to the dyads background information about volcanoes and a simulation displaying the sequence of seismic events that lead to volcanic eruptions. Dyads were then asked a series of

seven selected response questions in order to assess their knowledge of the previously presented information. Each participant was first prompted to respond to the question individually on the screen. System prompts were then displayed to facilitate collaborative dialogue as participants used a chat box to discuss their answers. Participants were next given the opportunity to revise the individual answer previously chosen and then one participant was randomly selected to submit a team response. When participants submitted the team response, their partners were able to see the selected answer (see Liu et al., 2016, for more in-depth discussion of the task). All of the turn-by-turn conversations and time-stamped responses to the questions were recorded in a carefully designed log file (Hao, Smith, Mislevy, von Davier, & Bauer, 2016). These conversations were used to measure CPS skills, and detect student roles using the GCA, while the responses to the in-simulation science items were used to measure science inquiry skills (Zapata-Rivera et al., 2014).

### 2.3. Quantifying CPS Skills

The chat communications during the collaboration were annotated in the previous study (Hao, Liu, von Davier, & Kyllonen, 2017) based on CPS framework developed by Liu and colleagues (2016). In their study, two human raters were trained on the CPS framework, and then independently coded the data. The unit of analysis was each turn of a conversation, or each conversational utterance. They reported an unweighted kappa of 0.65 for the four major categories, which is an acceptable level of agreement for social science experiments. The goal of this previous research using this dataset was to qualitatively assess learners' CPS skills, and understand their relationship to individual learner and group performance. In the current research, we focus on the four main CPS categories. Specifically, for each student we calculated the proportional occurrence for each of the four categories of CPS skills (e.g., share ideas, negotiate ideas, regulate problem-solving, and maintain communication). It is important to note that the computationally derived GCA measures (discussed below in section 2.5) were constructed completely independently of the qualitative CPS skill coding assessment, which was conducted in the previous study by Hao and colleagues (2017).

### 2.4. Quantifying CPS Outcomes

As discussed earlier, participants submitted both an individual answer, a revised individual answer, and a team answer to each of the 7 items. Student responses to the seven multiple-choice equivalent items were scored based on the corresponding scoring rubrics as presented in Zapata-Rivera et al. (2014). In this way, the responses before collaboration capture each individual member's science inquiry skills specific to the task, while the changes in individual student responses after the collaboration and the team responses reflect how effective the collaboration was and allow us to investigate which sociocognitive interaction dynamics may be more important for better collaboration skills and outcomes. Although these variables are correlated, they still provide a measure of how the interaction influences individual and team outcomes. Thus, in the current research we focus on the total correct score for individual student revised responses and the team response to quantify individual student and team level outcomes.

### 2.5. GCA Analysis Details

Five of GCA measures are semantic-based metrics (i.e., all but participation). The GCA relies on latent semantic analysis (LSA) to infer the semantic relationship among the individual contributions. LSA, an automated high-dimensional associative analysis of semantic structure in discourse, can be used to model and quantify the quality of coherence by measuring the semantic similarity of one section of text to the next. 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, 2013). When used to model discourse cohesion, LSA tracks the overlap and transitions of meaning of text segments throughout the discourse.

### 2.6. Statistical Analyses

Means were first calculated for each learner across the six GCA measures in order to provide global accounts of learner interactions, and the data was centred and normalized. Specifically, the normalization procedure involved Winsorizing the data based on each variable's upper and lower percentile. A cluster analysis was adopted to evaluate Research Question 1. Prior to clustering, multicollinearity and collinearity were assessed through variance inflation factor (VIF), and Pearson correlations, respectively. The results support the view that collinearity and multicollinearity were not an issue with  $r < .7$  and  $VIF < 3$  (Fox & Weisberg, 2010). Given that cluster analysis can return any number of specified clusters, we used the NbClust R package, which provides 26 indices for determining the relevant number of clusters (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). Detailed specifications of each index can be found in Charrad et al. (2014), and majority vote between indices indicated  $k = 5$  was appropriate. Partitioning Around Medoids (PAM) was used to identify the communication patterns associated with specific learner clusters or roles during CPS interactions.

For research questions 1 and 2, a mixed-effects modelling methodology was adopted (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 modelling handles the hierarchical nesting, interdependency, and unit of analysis problems inherent in collaborative learning data. They are the most appropriate technique for investigating data derived from group interactions (De Wever, Van Keer, Schellens, & Valcke, 2007; Janssen, Erkens, Kirschner, & Kanselaar, 2011).

In addition to constructing the fixed effects models, null models with the random effects (the student nested in the teams or the team) but no fixed effects were also constructed. A comparison of the null random-effects-only model with the fixed-effect models allows us to determine whether social roles and communication patterns predict student and team performance above and beyond the variance attributed to individual students or teams. The Akaike Information Criterion (AIC), Log Likelihood (LL), and a likelihood ratio test were all used to evaluate the overall fit 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 in R (Pinheiro et al., 2016) was used to perform all the required computations.

### 3. Results and Discussion

#### 3.1. Detecting Sociocognitive Roles

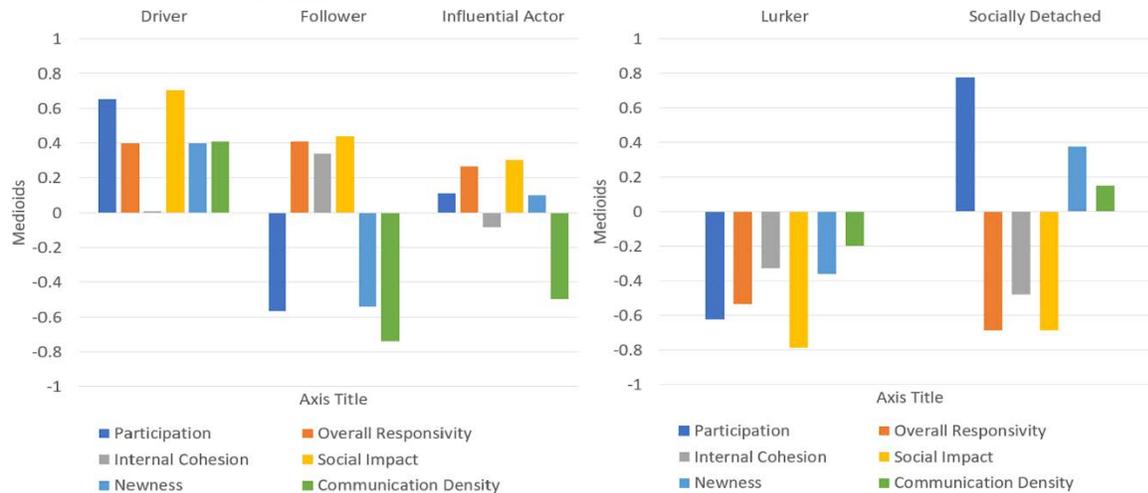
A cluster analysis approach was adopted to discover communication patterns associated with specific learner roles during collaborative problem-solving. Cluster analysis has been applied in previous studies of social roles (e.g., Lehmann-Willenbrock, Beck, & Kauffeld, 2016; Risser & Bottoms, 2014) and has proven useful in building an understanding of individual behaviours in many digital environments more broadly (Hou, 2015; Lankton, McKnight, & Tripp, 2017; Mirriahi, Liaqat, Dawson, & Gašević, 2016; Wise, Speer, Marbouti, & Hsiao, 2012). Analyses were carried out as described in section 2.6. Descriptive statistics for the GCA measures (z-scores) are presented in Table 2.

**Table 2. Descriptive Statistics for GCA Measures**

Measure	Minimum	Media n	$M$	$SD$	Maximum
Participation	-0.25	0	0	0.06	0.25
Social Impact	0.03	0.1	0.1	0.03	0.3
Overall Responsivity	0.03	0.1	0.1	0.03	0.3
Newness	0.12	0.49	0.49	0.1	1.34
Internal Cohesion	0.03	0.1	0.11	0.04	0.36
Communication Density	0.06	0.19	0.19	0.06	0.55

*Note:* Mean ( $M$ ); Standard deviation ( $SD$ ).

A first step in interpreting the clusters involves inspecting the cluster medoids, as this sheds light on whether the segments are conceptually distinguishable. Medoids are representative objects, or in this context learners, of a cluster whose average dissimilarity to all the other learners in the cluster is minimal. In contrast to the centroids used in the k-means algorithm, the medoids from PAM are represented by actual data points that best characterize the cluster. The medoids for the five-cluster PAM solution are presented below in Figure 2. It is worth noting that since the clustering was performed on normalized data, zero in this figure represents the population average for each measure, while positive and negative values represent values above or below that average, respectively.



**Figure 2. Medoids for the five-cluster solution across the GCA variables. More socially engaged roles (left) and less socially engaged roles (right).**

We see some interesting patterns across the five-cluster solution. The learners of the first three clusters take on more socially engaged roles. In line with Dowell, Nixon, and Graesser (2018), they are labelled **Drivers** ( $N = 168$ ), **Followers** ( $N = 181$ ), and **Influential Actors** ( $N = 173$ ). **Drivers** scored among the highest in all measures except internal cohesion. These individuals show investment in a high degree of effort in collaborative discussion displaying self-regulatory and social regulatory skills. **Followers** have low participation, but scored high in internal cohesion, overall responsiveness, and social impact. This pattern is suggestive of a student engaged in collaborative problem-solving, but perhaps taking a more thoughtful and deliberative stance than a **Driver**. **Influential Actors** contribute just slightly above average; however, their participation appears to build, at least minimally, on previously contributed ideas and move the collaboration forward (i.e., moderately positive social impact, responsiveness, and newness).

The learners of the final two clusters appear to take on less socially engaged roles than those of the previous three. These roles are labelled **Lurkers** ( $N = 242$ ) and **Socially Detached** ( $N = 203$ ). **Lurkers** have some of the lowest values across all GCA measures. In previous literature, lurkers have been defined differently, ranging from non-participants to minimal participants (Preece, Nonnecke, & Andrews, 2004). Lurkers’ low scores across the GCA measures do not suggest a complete lack of social impact or responsiveness to others. Rather, it suggests that these students expressed far less compared to the population average. **Socially Detached** learners have the highest participation among any of the clusters and score quite high in communication density and newness. However, they score very low across measures of responsivity suggesting that their discourse is more in response to themselves than others since they exhibit relatively higher internal cohesion compared to overall responsiveness. Overall, the five-cluster model appears, at least upon an initial visual inspection, to produce theoretically meaningful participant roles.

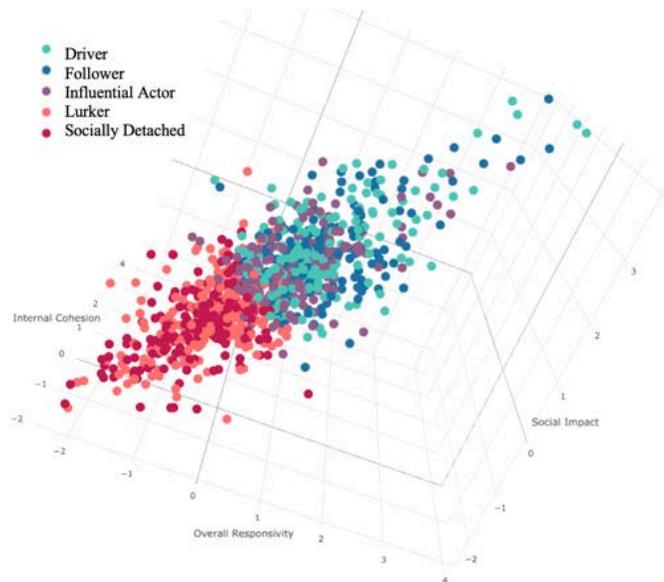
### 3.2. Clustering Evaluation and Validation

The literature proposes 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 the data is better than an alternative (Tanir, 2006). In the current research, we evaluated the internal, cluster coherence, and stability validation criteria. A commonly reported internal validity measure, Silhouette, measures how well an observation is clustered by estimating the average distance between clusters and ranges from  $-1$  (indicating that observations are likely placed in the wrong cluster) to  $1$  (indicating that the clusters perfectly separate).

In the context of real-world data (i.e., not simulated data) silhouette coefficient values above zero are described as reflecting intermediate (Subbalakshmi, Krishna, Rao, & Rao, 2015) and values between  $.20$ – $.50$  are considered to have fair internal validity (Mooi & Sarstedt, 2011) for a cluster solution and are commonly reported in the literature (e.g., Postareff, Mattsson, Lindblom-Ylänne, & Hailikari, 2017). The average silhouette (AS) for our model was fair ( $AS = .24$ ), indicating that the learners in a cluster had higher similarity to other learners in their own cluster than to students in any other cluster.

An evaluation of cluster coherence can help to establish that the identified clusters do in fact represent distinct modes in the distribution of the GCA measures. A MANOVA, ANOVAs, followed by Tukey’s post hoc were used to evaluate cluster

coherence or the extent to which learners in the cluster groups significantly differed from each other on the six GCA variables. This coherence evaluation showed that the five-cluster model exhibited nice separation across the GCA measures (for details on these analyses, please see the supplementary material in Appendix A). Figure 3 provides an illustrative example of how groups of clusters divide along three of the GCA measures: social impact, responsivity, and internal cohesion. Here, we can see how the Driver, Follower, and Influential Actor exhibit significantly higher scores, and a larger dispersion along these sociocognitive variables, compared to Lurker and Socially Detached learners.



**Figure 3.** Illustrative example showing how clusters separate along sociocognitive responsivity measures, with the Driver, Follower, and Influential Actor exhibiting higher responsivity scores, compared to Lurker and Socially Detached learners.

The stability and validity of the cluster model was assessed using a non-parametric bootstrapping procedure ( $B=100$  runs), which resampled from the original data with replacement to construct bootstrap matrices and clusters, and iteratively used the Jaccard coefficient to compute the structural similarity of the resampled clusters with the cluster derived from the original data. The Jaccard's similarity values showed average to strong prediction for all five clusters with .67, .73, .61, .60, and .70 for clusters 1–5, respectively. Given the extent of these evaluations, we feel that the identified learner profiles can be considered as robust and stable constructs in the space of the current collaborative problem-solving interactions, and that the GCA measures capture the critical sociocognitive processes necessary for identifying such profiles.

## 4. Sociocognitive Roles and CPS Outcomes

In the current research, the usefulness of identifying learners' roles in collaborative problem-solving teams is explored through two analyses of the data: 1) the influence of student roles on individual student problem-solving performance, and 2) the influence of team composition properties on overall team performance (research question 2).

### 4.1. Influence of Individual Student Roles on Individual Student and Team CPS outcomes

In order to evaluate the effects of roles purely at the individual level, two linear mixed-effects models were compared: 1) a model with student performance as the dependent variable, social roles as independent variables, and student nested within the team as the random effects, and 2) the null model with random effects only and no fixed effects. We also assessed the potential influence of prior knowledge by including it as a covariate; however, this did not result in any significant interactions and, as such, it was removed from the model.

The likelihood ratio tests indicated that the five-role model with  $\chi^2(4) = 14.87$ ,  $p < .01$ ,  $R^2_m = .02$ ,  $R^2_c = .86$  yielded a significantly better fit than the null model. A number of conclusions can be drawn from these statistics. First, the roles in the five-cluster model were able to add significantly to the prediction of the learners' performance beyond the variance attributed to student and team 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 predictive of individual student learning performance for the five-cluster model are presented in Table 3. The reference group was the Influential Actor role, meaning that the performance for the other roles are compared against the Influential Actor reference group. Influential Actors were chosen as the reference group because they were among the highest performers; as such, they afforded statistical comparison that most directly addressed the substantive research questions. Four of the five roles exhibit significant differences in student performance, as compared to the Influential Actor role. Note that, as the reference group, the values for Influential Actors are evaluated to determine if they differ significantly from zero. Here we see learners who took on more socially responsible, collaborative roles, such as the Influential Actor, performed significantly better than students who occupied the less socially engaged roles (Lurker and Socially Detached). There was no significant difference between the performances of the Driver and Influential Actor (and only marginally significant for the Follower), suggesting that these are the more successful roles in terms of student-level collaborative problem-solving outcomes.

**Table 3.** Mixed-Effects Model Coefficients for Predicting Individual level CPS Outcomes with Sociocognitive Roles

Measure	Individual Performance	
	$\beta$	SE
Driver	-0.07	0.10
Follower	-0.24*	0.10
Influential Actor	0.19**	0.07
Lurker	-0.31***	0.10
Socially Detached	-0.29**	0.10

Note: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; Fixed effect coefficient ( $\beta$ ); Standard error (SE);  $N = 967$ .

It is important to note that the observed difference in CPS performance across the social roles is not a result of the students simply being more prolific because the Followers performed almost on par with the Influential Actors though among the lower participators in the team. The profile for Socially Detached learners showed lower values for responsivity and social impact, compared to their internal cohesion scores. However, the Followers profile illustrated that when they did make contributions, it was quite responsive to the other group members (i.e., high overall responsivity), as well as being semantically connected with their 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, Influential Actors, and Followers, compared to the Socially Detached, beyond the surface level mechanism of simply participating often. Taken together, these results show that the identified roles are externally valid, not just due to a significant relationship to the external measure of learning, but also because we can make theoretically meaningful predictions from the role’s associated characteristic behaviours.

#### 4.2 Influence of Group Composition on Team CPS outcomes

The second set of analyses involved a more fine-grained investigation of the influence of (the proportional occurrence of) positive and negative roles on team performance. Particularly, we constructed a productive roles model with the proportional occurrence of Drivers and Influential Actors as the independent variable, and an unproductive roles model with the proportional occurrence of Socially Detached, Followers, and Lurkers as the independent variable. Null models were constructed for both analyses. Team performance was the dependent variable in both models.

The likelihood ratio tests indicated that both the productive roles model and the unproductive roles model yielded a significantly better fit than the null model with  $\chi^2(2) = 8.75, p < .01, R^2_m = .02, R^2_c = .89$  and  $\chi^2(3) = 9.24, p < .05, R^2_m = .02, R^2_c = .88$ , respectively. Here, we see that the proportional occurrence of productive and unproductive roles, were able to significantly improve predictions of the team performance above and beyond the variance attributed to group. Secondly, 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 4% of the variance being accounted for by the proportional occurrence of different social roles. Table 4 shows the social roles predictive of group performance for both the productive and the unproductive roles model.

As shown in Table 4, groups with greater proportions of learners who took on more socially responsible, collaborative roles (namely Driver and Influential Actor) performed significantly better than groups with greater proportions of less socially engaged roles (e.g., Socially Detached). These findings mirror the pattern observed for individual student learning and social roles, with one exception. The proportional occurrence of Lurkers on a team does not significantly influence the team outcome.

**Table 4. Descriptive Statistics for Group Performance across Five Roles and Mixed-Effects Model Coefficients for Predicting the Influence of Productive and Unproductive Roles on Group Performance**

Role	Productive Roles Model				Role	Unproductive Roles Model			
	<i>M</i>	<i>SD</i>	$\beta$	<i>SE</i>		<i>M</i>	<i>SD</i>	$\beta$	<i>SE</i>
Prop. of Drivers	0.12	1.00	0.39*	0.18	Prop. of Followers	0.01	1.06	-0.51*	0.26
Prop. of Influential Actors	0.16	0.87	0.38**	0.16	Prop. of Lurkers	-0.08	1.01	-0.34	0.24
					Prop. of Socially Detached	-0.13	1.01	-0.46*	0.23

Note:  $N = 477$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; Mean (*M*); Standard deviation (*SD*); Fixed effect coefficient ( $\beta$ ); Standard error (*SE*).

### 5. Sociocognitive Roles and CPS Skills

The next research question addressed the associations between the identified roles and CPS skills. The question focused on how the roles and processes are associated with each other. To answer this question, four models were constructed to explore the relationship between CPS skills and the identified roles. Specifically, for each model, one of the four main CPS skills was the dependent variable (e.g., maintaining communication, sharing ideas, regulating problem solving, and negotiating ideas). Note, that this dependent variable represents the proportional frequency each student exhibited in each of the four CPS skills throughout the interaction. The five roles were the independent variable in all four models. In line with the previous analyses, the reference group was the Influential Actor role, meaning that the observed CPS skills for the other roles are compared against the Influential Actor.

The likelihood ratio tests indicated that the **Maintaining Communication, Sharing Ideas, Regulating Problem Solving, and Negotiating Ideas** models all yielded a significantly better fit than their null model with  $\chi^2(4) = 85.58, p < .001, R^2_m = .10, R^2_c = 1.00, \chi^2(4) = 56.67, p < .001, R^2_m = .06, R^2_c = 1.00, \chi^2(4) = 22.05, p < .001, R^2_m = .02, R^2_c = 1.00,$  and  $\chi^2(4) = 90.27, p < .001, R^2_m = .09, R^2_c = 1.00,$  respectively. Here, we see that the roles were able to significantly improve predictions of the CPS skills above and beyond the variance attributed to individual students and teams. Secondly, for all models, the social roles combined with student and team features explained about 100% of the predictable variance in group performance, with 27% of the variance accounted for by social roles. Table 5 shows the social roles that were predictive of the four main CPS skills.

**Table 5. Mixed-Effects Model Coefficients for Predicting CPS Skills with Sociocognitive Roles**

Role	Maintaining Communication		Sharing Ideas		Regulating Problem Solving		Negotiating Ideas	
	$\beta$	<i>SE</i>	$\beta$	<i>SE</i>	$\beta$	<i>SE</i>	$\beta$	<i>SE</i>
Driver	-0.08	0.07	-0.04	0.10	-0.28**	0.10	0.36**	0.10
Follower	-0.11	0.07	-0.35***	0.10	-0.44***	0.10	0.77***	0.10
Influential Actor	-0.22***	0.06	0.33***	0.07	0.23**	0.08	-0.25**	0.07
Lurker	0.56***	0.08	-0.61***	0.10	-0.24*	0.10	0.16	0.09
Socially Detached	0.58***	0.08	-0.48***	0.10	-0.14	0.10	-0.06	0.10

Note: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ; Fixed effect coefficient ( $\beta$ ); Standard error (*SE*);  $N = 967$ .

From the previous investigation, we know that the more productive roles, in terms of CPS outcomes, are the Influential Actors, Drivers, and, to a lesser extent, Followers. As such, we use this information to understand the relative importance of CPS skills and further contextualize the roles. The first finding to note is that, across the four CPS skills, a relatively consistent overall pattern emerged (ignoring significance). Specifically, we observed a consistent pattern within less productive roles, Socially Detached learners and Lurkers, wherein they both engaged more in maintaining communication and less in sharing

ideas and regulating problem solving. This pattern deviates from the consistent pattern seen across the more productive roles, Drivers and Followers, that showed increased negotiating ideas coupled with less maintaining communication, regulating problem solving, and sharing ideas. Note that the Influential Actors deviate from this pattern as they are the reference group within the model. Further, the overall trend of the relationship, as compared to Influential Actors, plays out differently for more and less productive role groups.

Maintaining Communication is described as content irrelevant communication (Liu et al., 2016). In Table 5, we see that the less productive roles exhibited significantly more task irrelevant information compared to the Influential Actors, who exhibited significantly less. Further, Drivers and Followers were on par with the Influential Actors for Maintaining Communication. Conversations around more surface level, but task relevant information (e.g., student responds to teammate's request; student points out resource) are captured by the CPS Sharing Information skill. The results here show that the discourse of Influential Actors and Drivers was characterized by higher levels of sharing information, compared to the other roles. Conversations around goal setting, progress monitoring, and reflecting are all captured within the Regulating Problem Solving CPS skill. Influential Actors and Socially Detached learners demonstrated higher levels of problem-solving regulation during the team interactions, compared the other roles. Based on the distinct differences observed for the cluster profiles of these groups, we can discern that the nature of how they engaged in problem regulation differed, with Socially Detached learners engaging in a more personally oriented reflection, goal setting, and progress monitoring compared to their more team-oriented Influential Actor peers.

Negotiating ideas includes communication about the team members' collaborative knowledge building and construction through comparing alternative ideas and presenting evidence and rationale for an argument. In previous investigations, negotiating ideas was found to be the strongest differentiator between effective and ineffective CPS teams (Hao, Liu, von Davier, & Kyllonen, 2017; Hao et al., 2017; von Davier et al., 2017). Interestingly, the Influential Actors exhibited significantly lower negotiation discourse, and were on par with the less productive roles, namely Lurkers and Socially Detached learners. In contrast, the negotiation of ideas appears to be a skill where Drivers and Followers excelled.

## 6. General Discussion and Conclusion

Collaborative problem-solving (CPS) has become an essential component of today's knowledge-based, innovation-centred economy and society. As such, communication and CPS are now considered critical 21st century skills and have taken a principal role in educational practice, policy, and research. A primary objective of this research was to evaluate the relationship between learners' sociocognitive roles with CPS skills and outcomes. Despite general agreement that communication and CPS are important skills, methodological approaches for automated assessment of these vital sociocognitive processes during team interactions have remained limited. To address this, we used a new methodological approach, Group Communication Analysis (GCA), in the context of collaborative problem-solving interactions. There are multiple lenses by which one might view discourse and learner profiles in technology-mediated learning environments (e.g., Chua, Tagg, Sharples, & Rienties, 2017; Hecking, Chounta, & Hoppe, 2016; Kizilcec, Piech, & Schneider, 2013). The GCA lets us view discourse as a dynamic and evolving sociocognitive process that resides in the interaction between learners' communicative contributions. Our results suggest that these sociocognitive discourse patterns diagnostically reveal learners' interaction profiles, and that the observed patterns are robust.

The study presents the interpretation of learners' emergent sociocognitive roles in CPS and the methodological validation of such profiling. Here, we observed that participants exhibited a dominant role, which emerged and was shaped through consistent patterns in their discourse. Specifically, in section 3.1 (Detecting Sociocognitive Roles), we identified that the roles participants occupy while engaging in problem-solving activities have distinct patterns in behavioural engagement style (i.e., active or passive, leading or following), semantic characteristics (i.e., providing new information or echoing given material, ignoring or responding to teammates), and social orientation. The identified social roles (i.e., clusters) underwent stringent evaluation, and validation assessments to be considered as robust constructs. Interestingly, the GCA revealed roles that do not entirely overlap with those observed in previous research efforts (e.g., Dowell, Poquet, & Brooks, 2018; Strijbos & Weinberger, 2010). The identification of these new roles might serve as a conceptual basis for future research focusing on understanding the social roles within collaborative problem-solving communication.

We investigated the practical value of the identified roles in section 4 (Sociocognitive Roles and CPS outcomes), and whether they were meaningfully related to learner and team performance in CPS environments. Overall, the results suggest that the roles learners occupy while engaging in problem-solving activities influence their problem-solving efficacy, and that the pattern of roles that emerge within a team can be either more or less beneficial for the overall team's outcome. This analysis yielded an important contribution to the CPS literature. Some of the most noteworthy discoveries concern the influence of roles on student learning and team performance, which suggests that the difference in outcome measures across the roles is not a result of individuals simply being more active in terms of a number of contributions. The quality of conversation, as captured

by the internal cohesion, responsivity, and social impact measures, more than the quantity, appears to be the key element in the success for both teams and individuals during problem-solving activities. This is further manifested through the performance associated with productive versus unproductive roles.

Research in CPS has historically focused on assessing important aspects of CPS such as individual behaviours, communication, and collaboration competencies. These frames prioritize research detailing the characteristics and behaviours of interacting individuals, while the sociocognitive identity concepts that lead to individual CPS competencies have received far less attention. In section 5 (Sociocognitive Roles and CPS skills), we investigated the relationship between participants' individual CPS skills and their roles. Here, we observed a strong relationship between the roles emergent learners occupied and the skills they utilized during problem-solving. A noteworthy finding here pertains to the overall pattern across the CPS skills in relation to learner role. Specifically, we observed a consistent pattern within less productive roles — Socially Detached learners and Lurkers — that is distinctly different from the consistent pattern seen across the more productive roles — Drivers and Followers. Further, the overall trend of the relationship, as compared to Influential Actors, plays out differently for more and less productive role groups. It appears that the differences in the exhibited CPS skills between more and less productive roles reside in the fact that less productive roles engaged in more surface level CPS techniques, such as maintaining communication, and less constructive CPS discourse compared to Influential Actors, such as sharing ideas. Contrastively, more productive roles appeared to take a different path wherein they were on par with Influential Actors in minimal engagement with surface level CPS skills (i.e., maintaining communication), but diverged from Influential Actors regarding more substantive CPS engagement by exhibiting less regulating and problem-solving behaviour and more negotiating behaviour. This suggests that perhaps the students occupying less productive roles could benefit from support in regulation, such as socially shared motivation regulation (Järvelä, Järvenoja, Malmberg, & Hadwin, 2013).

The contributions from this line of research are twofold. First, as a conceptual contribution, we have gained additional insights regarding how individuals personalize their CPS skill competency efforts, and how these are constructed and maintained across metalevel roles and micro-level individual competencies. This information can also be seen as a methodological contribution, as we have provided a means for triangulating qualitative analysis of individual learners' CPS competencies with automated and scalable quantitative analysis of user roles in the sociocognitive and semantic context of CPS interactions. Additionally, triangulating GCA with qualitative analysis provides additional validation of the identified roles and sociocognitive processes within CPS. As another methodological contribution, the classification of types of CPS skills and behaviours into a smaller set of sociocognitive roles reduces the analytic complexity of capturing CPS skills within a social system and facilitates the comparative study of populations across time and setting. As a more practical contribution, such information can now be easily accessible to educational stakeholders to improve the design of technology-mediated CPS environments. However, further validation of these profiles is needed, including replicating the results in other contexts and with other CPS skill frameworks.

Researchers previously expressed the need for more in-depth analysis for CPS interactions, as well as for adapting the growing body of discourse data in online collaborative environments (Care et al., 2016; Dowell, Nixon, & Graesser, 2018; Dowell, Poquet, & Brooks, 2018; Graesser, Cai, Morgan, & Wang, 2017; Rosen & Foltz, 2014; Scoular & Care, 2019; Scoular, Care, & Hesse, 2017; Wise & Schwarz, 2017). By pursuing a purely automated computational-linguistic methodology, we were able to explore a substantially larger number of problem-solving interactions than could typically be analyzed with manual content analysis. Indeed, one may argue that the efficiency may come at a cost of accuracy compared to human judgment. By recognizing that machine-learned method has its inherent limitations, we performed stringent tests of cluster stability, coherence, and internal consistency to eliminate those concerns. As we show that identified roles are indicative of both CPS performance and skills, the analysis is proven to be effective and meaningful in understanding multifaceted interactions.

This research serves as an evaluation of the GCA in the context of collaborative problem-solving and understanding the relationship between sociocognitive roles and problem-solving skills and outcomes. 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. Using machine learning models, GCA overcomes some of the constraints inherent in manual qualitative analysis or summative data streams and provides novel perspectives with quantified intra- and interpersonal dynamics involved in CPS. As a methodological contribution, we expect the GCA will provide a more objective, domain-independent means for future exploration of roles in CPS 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. With its automated processes, GCA has the capacity to provide real-time evaluation in large-scale STEM CPS environments efficiently and effectively. This offers a unique

opportunity to observe CPS processes as they unfold and build responsive systems that capture positive or negative interactions among learners. Instructors may make use of this to monitor live performance at both individual and group levels, react to signs of malfunctioning group behaviours from early on, and adapt teaching strategies accordingly to scaffold successful group collaborative problem-solving. Such analytics-based assessment may also be embedded in a learner-facing dashboard to inform learners about their roles in the group and their development of CPS skills, creating room for future experiment intervention to take a deeper investigation on self-regulated learning, motivation, and other group factors associated with learner outcomes. Overall, 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.

## Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The publication of this article received financial support from the National Science Foundation under Grant No. 1535300 and Grant No. BCC 14-517. This work was also supported in part by the Holistic Modelling of Education (HOME) grant from the Michigan Institute for Data Science (MIDAS). Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of these funding agencies.

## References

- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, 9(4), 403–436. [https://dx.doi.org/10.1207/S15327809JLS0904\\_2](https://dx.doi.org/10.1207/S15327809JLS0904_2)
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12(3), 307–359. [https://dx.doi.org/10.1207/S15327809JLS1203\\_1](https://dx.doi.org/10.1207/S15327809JLS1203_1)
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2011). Defining 21st century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching 21st century skills* (pp. 17–66). Heidelberg: Springer. [https://dx.doi.org/10.1007/978-94-007-2324-5\\_2](https://dx.doi.org/10.1007/978-94-007-2324-5_2)
- Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, 29(4), 250–264. <https://dx.doi.org/10.1080/08957347.2016.1209204>
- Cesareni, D., Cacciamani, S., & Fujita, N. (2015). Role taking and knowledge building in a blended university course. *International Journal of Computer-Supported Collaborative Learning*, 11(1), 9–39. <https://dx.doi.org/10.1007/s11412-015-9224-0>
- 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). <https://dx.doi.org/10.18637/jss.v061.i06>
- Chi, M. T. H. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. <https://dx.doi.org/10.1111/j.1756-8765.2008.01005.x>
- Chua, S. M., Tagg, C., Sharples, M., & Rienties, B. (2017). Discussion analytics: Identifying conversations and social learners in FutureLearn MOOCs. In L. Vigentini & M. L. Urrutia (Eds.), *Proceedings of the 7th International Conference on Learning Analytics and Knowledge (LAK '17)*, 13–17 March 2017, Vancouver, BC, Canada (pp. 36–62). New York: ACM. Retrieved from <http://oro.open.ac.uk/id/eprint/57071>
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, H. 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. <https://dx.doi.org/10.1037/10096-006>
- Dede, C. (2010). Comparing frameworks for 21st century skills. In J. Bellanca & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 51–76). Bloomington, IN: Solution Tree Press.
- 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(4), 436–447. <https://dx.doi.org/10.1016/j.learninstruc.2007.04.001>
- Dillenbourg, P., & Fischer, F. (2007). Basics of computer-supported collaborative learning. *Zeitschrift Fur Berufs-Und Wirtschaftspädagogik*, 21, 111–130.
- Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-enhanced learning: Principles and products* (pp. 3–19). Dordrecht: Springer Netherlands. [https://dx.doi.org/10.1007/978-1-4020-9827-7\\_1](https://dx.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(1), 121–151. [https://dx.doi.org/10.1207/s15327809jls1501\\_9](https://dx.doi.org/10.1207/s15327809jls1501_9)
- Doise, W. (1990). The development of individual competencies through social interaction. In H. C. Foot, M. Morgan, & R. H. Shute (Eds.), *Children helping children*. Chichester, UK: J. Wiley & Sons.
- Dornfeld, C., & Puntambekar, S. (2015). Emergent roles and collaborative discourse over time. In O. Lindwall, P. Hakkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden (pp. 380–387). International Society of the Learning Sciences. Retrieved from <https://www.isls.org/cscl2015/papers/MC-0347-FullPaper-Dornfeld.pdf>
- Dowell, N. M. (2017). A computational linguistic analysis of learners' discourse in computer-mediated group learning environments. PhD Dissertation, University of Memphis.
- Dowell, N. (2019). Preparing for the future: Group communication analysis as a tool to facilitate adaptive support during digitally-mediated team interactions. *Proceedings of the 3rd International Conference on AI + Adaptive Education*, (AIAED). Beijing, China. New York: ACM.
- Dowell, N. M., Lin, Y., Godfrey, A., Cho, H., & Brooks, C. (2019). Promoting inclusivity through time-dynamic discourse analysis in digitally-mediated collaborative learning. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Proceedings of the 20th International Conference on Artificial Intelligence in Education (AIED 2019)*, 25–29 June 2019, Chicago, IL, USA (pp. 207–219). Springer. [https://dx.doi.org/10.1007/978-3-030-23204-7\\_18](https://dx.doi.org/10.1007/978-3-030-23204-7_18)
- Dowell, N. M., Nixon, T., & Graesser, A. C. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multi-party interactions. *Behavior Research Methods*, 51(3), 1007–1041. <https://dx.doi.org/10.3758/s13428-018-1102-z>
- Dowell, N. M., Poquet, O., & Brooks, C. (2018). Applying group communication analysis to educational discourse interactions at scale. In J. Kay & R. Luckin (Eds.), *Rethinking Learning in the Digital Age: Making the Learning Sciences Count. Proceedings of the 13th International Conference of the Learning Sciences (ICLS '18)*, 23–27 June 2018, London, UK (3 volumes, pp. 1815–1822). International Society of the Learning Sciences. <https://dx.doi.org/10.22318/cscl2018.1815>
- Fiore, S. M., Graesser, A., & Greiff, S. (2018). Collaborative problem-solving education for the twenty-first-century workforce. *Nature Human Behaviour*, 2(6), 367–369. <https://dx.doi.org/10.1038/s41562-018-0363-y>
- Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(2), 203–224. <https://dx.doi.org/10.1177/0018720810369807>
- Fiore, S. M., & Schooler, J. W. (2004). Process mapping and shared cognition: Teamwork and the development of shared problem models. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 133–152). Washington, DC: American Psychological Association. <https://dx.doi.org/10.1037/10690-007>
- Fox, J., & Weisberg, H. S. (2010). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415. <https://dx.doi.org/10.1073/pnas.1319030111>
- Gervits, F., Eberhard, K., & Scheutz, M. (2016). Team communication as a collaborative process. *Frontiers in Robotics and AI*, 3, 62. <https://dx.doi.org/10.3389/frobt.2016.00062>
- Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). A conceptual and operational definition of “social role” in online community. *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS-42)*, 5–8 January 2009, Waikoloa, Big Island, HI, USA (pp. 1–11). IEEE Computer Society. <https://dx.doi.org/10.1109/HICSS.2009.6>
- Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*, 76, 607–616. <https://dx.doi.org/10.1016/j.chb.2017.03.041>
- Graesser, A. C., Dowell, N., & Clewley, D. (2017). Assessing collaborative problem solving through conversational agents. In A. A. Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 65–80). Springer. [https://dx.doi.org/10.1007/978-3-319-33261-1\\_5](https://dx.doi.org/10.1007/978-3-319-33261-1_5)
- Graesser, A. C., Dowell, N. M., Hampton, A., Lippert, A. M., Li, H., & Shaffer, D. W. (2018a). Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 program for

- international student assessment. In J. J. Johnston, R. Sottolare, A. Sinatra, & C. S. Burke (Eds.), *Building intelligent tutoring systems for teams: What matters*. (Vol. 19, pp. 173–214). West Yorkshire, UK: Emerald Publishing.
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018b). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, 19(2), 59–92. <https://dx.doi.org/10.1177/1529100618808244>
- Graesser, A. C., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. L. (2018c). Challenges of assessing collaborative problem solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills* (pp. 75–91). Springer, Cham. [http://dx.doi.org/10.1007/978-3-319-65368-6\\_5](http://dx.doi.org/10.1007/978-3-319-65368-6_5)
- Greenhow, C., Robelia, B., & Hughes, J. E. (2009). Learning, teaching, and scholarship in a digital age: Web 2.0 and classroom research: What path should we take now? *Educational Researcher*, 38(4), 246–259.
- Guo, Z., Yu, K., Pearlman, R., Navab, N., & Barmaki, R. (2019). Collaboration analysis using deep learning. arXiv [cs.HC]. Retrieved from <http://arxiv.org/abs/1904.08066>
- Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 135–156). Springer, Cham. [https://dx.doi.org/10.1007/978-3-319-33261-1\\_9](https://dx.doi.org/10.1007/978-3-319-33261-1_9)
- Hao, J., Liu, L., von Davier, A. A., Lederer, N., Zapata-Rivera, D., Jakl, P., & Bakkenson, M. (2017). EPCAL: ETS platform for collaborative assessment and learning. ETS Research Report Series, 2017(1), 1–14. <https://dx.doi.org/10.1002/ets2.12181>
- Hao, J., Liu, L., von Davier, A., & Kyllonen, P. (2015). Assessing collaborative problem solving with simulation based tasks. In O. Lindwall, P. Hakkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015)*, 7–11 June 2015, Gothenburg, Sweden (pp. 544–547). International Society of the Learning Sciences. Retrieved from <https://pdfs.semanticscholar.org/c922/044e125a053be368e3eb146c9cf581fe55c5.pdf>
- Hao, J., Smith, L., Mislavy, R., von Davier, A., & Bauer, M. (2016). Taming log files from game/simulation-based assessments: Data models and data analysis tools. ETS Research Report Series, 2016(1), 1–17. <https://dx.doi.org/10.1002/ets2.12096>
- 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: Oxford University Press.
- Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK '16)*, 25–29 April 2016, Edinburgh, UK (pp. 198–207). New York: ACM. <https://dx.doi.org/10.1145/2883851.2883924>
- Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of cluster analysis*. New York: CRC Press. <https://dx.doi.org/10.1201/b19706>
- 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). Springer Netherlands. Retrieved from [http://link.springer.com/chapter/10.1007/978-94-017-9395-7\\_2](http://link.springer.com/chapter/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(6), 571–606. <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(1), 48–94. <http://dx.doi.org/10.1080/07370000701798495>
- Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, 48, 424–435. <https://dx.doi.org/10.1016/j.chb.2015.02.010>
- Howley, I. K., & Rosé, C. P. (2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262. <https://dx.doi.org/10.18608/jla.2016.33.12>
- Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, 51(4), 1755–1765. <https://dx.doi.org/10.1016/j.compedu.2008.05.005>
- Hu, X., Dowell, N., Cai, Z., Graesser, A. C., Shi, G., Cockroft, J., & Shorter, P. (2018). Constructing individual conversation characteristics curves (ICCC) for interactive intelligent tutoring environments (IITE). In R. Sottolare, A. Graesser, X. Hu, & A. Sinatra (Eds.), *Design recommendations for intelligent tutoring systems, Volume 6: Team cognition*. Orlando, FL: US Army Research Laboratory.
- 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). Boston, MA: Springer US. [https://dx.doi.org/10.1007/978-1-4419-7710-6\\_9](https://dx.doi.org/10.1007/978-1-4419-7710-6_9)
- Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. *Journal of Cognitive Education and Psychology*, 12, 267–286. <http://dx.doi.org/10.1891/1945-8959.12.3.267>
- Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive–situative divide and combining individual and social processes. *Educational Psychologist*, 45(1), 15–27. <https://dx.doi.org/10.1080/00461520903433539>
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, 51(2), 247–265. <http://dx.doi.org/10.1080/00461520.2016.1158654>
- 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(2), 403–420. <https://dx.doi.org/10.1016/j.chb.2007.01.028>
- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 5–10 May 2008, Florence, Italy (pp. 453–456). New York: ACM. <https://dx.doi.org/10.1145/1357054.1357127>
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK '13)*, 8–12 April 2013, Leuven, Belgium (pp. 170–179). New York: ACM. <https://dx.doi.org/10.1145/2460296.2460330>
- Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground coordination in joint activity. In W. B. Rouse & K. R. Boff (Eds.), *Organizational simulation* (pp. 139–184). John Wiley & Sons. <https://dx.doi.org/10.1002/0471739448.ch6>
- 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(3), 335–353. [https://dx.doi.org/10.1016/S0747-5632\(02\)00057-2](https://dx.doi.org/10.1016/S0747-5632(02)00057-2)
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, 44(1), 46–53. <https://dx.doi.org/10.3102/0013189X15569530>
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013). *Handbook of latent semantic analysis*. New York: Routledge.
- Lankton, N. K., McKnight, D. H., & Tripp, J. F. (2017). Facebook privacy management strategies: A cluster analysis of user privacy behaviors. *Computers in Human Behavior*, 76, 149–163. <https://doi.org/10.1016/j.chb.2017.07.015>
- Lehmann-Willenbrock, N., Beck, S. J., & Kauffeld, S. (2016). Emergent team roles in organizational meetings: Identifying communication patterns via cluster analysis. *Communication Studies*, 67(1), 37–57. <https://dx.doi.org/10.1080/10510974.2015.1074087>
- Letsky, M. P. (2008). *Macrocognition in teams: Theories and methodologies*. Aldershot, UK: Ashgate.
- Lin, Y., Dowell, N. M., Godfrey, A., Cho, H., & Brooks, C. (2019). Modeling gender differences in intra- and interpersonal dynamics during collaborative interactions. In D. Azcona & R. Chung (Eds.), *Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK '19)*, 4–8 March 2019, Tempe, Arizona, USA (pp. 431–435). New York: ACM.
- Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J.-D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global. <https://dx.doi.org/10.4018/978-1-4666-9441-5.ch013>
- Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, 49(Supplement C), 160–174. <https://dx.doi.org/10.1016/j.cedpsych.2017.01.009>
- Mesmer-Magnus, J. R., & Dechurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94(2), 535–546. <https://dx.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. <https://dx.doi.org/10.1007/s11423-016-9449-2>
- Mooi, E., & Sarstedt, M. (2011). Cluster analysis. In *A concise guide to market research: The process, data, and methods using IBM SPSS statistics* (pp. 237–284). Springer. <https://dx.doi.org/10.1007/978-3-642-12541-6>

- 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(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- National Research Council. (2011). *Assessing 21st Century Skills: Summary of a Workshop*. Washington, DC: National Academies Press. <https://dx.doi.org/10.17226/13215>
- OECD. (2013). *PISA 2015 collaborative problem solving framework*. Oxford, UK: OECD Publishing.
- Piaget, J. (1993). *The moral judgement of the child*. New York: Simon & Schuster.
- Pinheiro, J., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. Springer. Retrieved from <http://books.google.co.in/books?id=N3WeyHFbHLQC>
- 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>
- Postareff, L., Mattsson, M., Lindblom-Ylänne, S., & Hailikari, T. (2017). The complex relationship between emotions, approaches to learning, study success and study progress during the transition to university. *Higher Education*, 73(3), 441–457. <https://dx.doi.org/10.1007/s10734-016-0096-7>
- Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior*, 20(2), 201–223. <https://dx.doi.org/10.1016/j.chb.2003.10.015>
- Price, E. F. (1981). Toward a taxonomy of given/new information. In P. Cole (Ed.), *Radical pragmatics*. New York: 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(4), 433–450. <https://dx.doi.org/10.1007/s11412-014-9197-4>
- Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, 2(3), 235–276. [https://dx.doi.org/10.1207/s15327809jls0203\\_1](https://dx.doi.org/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: Springer-Verlag. [http://dx.doi.org/10.1007/978-3-642-85098-1\\_5](http://dx.doi.org/10.1007/978-3-642-85098-1_5)
- Rosé, C. P., Stahl, G., Goggins, S., Patterson, E., Borge, M., Carroll, J., & Duchon, A. (2014). *Towards optimization of macrocognitive processes: Automating analysis of the emergence of leadership in ad hoc teams*. Pittsburgh, PA: Carnegie-Mellon University.
- Rosen, Y., & Foltz, P. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning*, 9(3), 389–410. <http://dx.doi.org/10.4018/978-1-4666-5888-2.ch010>
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18689065>
- Scoular, C., & Care, E. (2019). Monitoring patterns of social and cognitive student behaviors in online collaborative problem solving assessments. *Computers in Human Behavior*, 104. <https://dx.doi.org/10.1016/j.chb.2019.01.007>
- Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment: Operationalizing collaborative problem solving. *Journal of Educational Measurement*, 54(1), 12–35. <https://dx.doi.org/10.1111/jedm.12130>
- Siddiq, F., & Scherer, R. (2017). Revealing the processes of students’ interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior*, 76, 509–525. <https://dx.doi.org/10.1016/j.chb.2017.08.007>
- Singley, M. K., Fairweather, P. G., & Swerling, S. (1999). Team tutoring systems: Reifying roles in problem solving. In C. Hoadley & J. Roschelle (Eds.), *Proceedings of the 1999 Conference on Computer-Supported Collaborative Learning (CSCL ’99)*, 12–15 December 1999, Palo Alto, CA, USA (pp. 66–es). International Society of the Learning Sciences. Retrieved from <http://dl.acm.org/citation.cfm?id=1150240.1150306>
- Spada, H. (2010). Of scripts, roles, positions, and models. *Computers in Human Behavior*, 26(4), 547–550. <https://dx.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). Hillsdale, NJ: Lawrence Erlbaum. Retrieved from [https://pdfs.semanticscholar.org/e717/72b208cdc16c2851eaa287e58be9bec19d22.pdf?\\_ga=2.79495700.1684383047.1580443850-1453764810.1580443850](https://pdfs.semanticscholar.org/e717/72b208cdc16c2851eaa287e58be9bec19d22.pdf?_ga=2.79495700.1684383047.1580443850-1453764810.1580443850)
- 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). London: Routledge.
- Strijbos, J.-W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior*, 26(4), 491–494. <https://dx.doi.org/10.1016/j.chb.2009.08.006>

- Subbalakshmi, C., Krishna, G. R., Rao, S. K. M., & Rao, P. V. (2015). A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set. *Procedia Computer Science*, 46, 346–353. <https://dx.doi.org/10.1016/j.procs.2015.02.030>
- Suthers, D. 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(1), 5–42.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Taniar, T. (2006). *Research and trends in data mining technologies and applications*. Hershey, PA: IGI Global. Retrieved from <https://books.google.com/books?id=17bSnt3DGUYC>
- Teasley, S. D., Fischer, F., Weinberger, A., Stegmann, K., Dillenbourg, P., Kapur, M., & Chi, M. (2008). Cognitive convergence in collaborative learning. *Proceedings of the 8th International Conference on the Learning Sciences (ICLS '08)*, 24–28 June 2008, Utrecht, Netherlands (Vol. 3, pp. 630–637). International Society of the Learning Sciences.
- Teasley, S. D., & Roschelle, J. (1993). Constructing a joint problem space: The computer as a tool for sharing knowledge. In S. P. Lajoie & S. L. Derry (Eds.), *Computers as cognitive tools* (pp. 229–258). Hillsdale, NJ: Erlbaum.
- Van Boxtel, C. (2004). Studying peer interaction from three perspectives. In J. Linden & P. Renshaw (Eds.), *Dialogic learning* (pp. 125–143). New York: Springer. Retrieved from [http://link.springer.com/chapter/10.1007/1-4020-1931-9\\_7](http://link.springer.com/chapter/10.1007/1-4020-1931-9_7)
- van den Bossche, P., 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(5), 490–521. Retrieved from <https://dspace.library.uu.nl/handle/1874/16978>
- von Davier, A. A., & Halpin, P. F. (2013). Collaborative problem solving and the assessment of cognitive skills: Psychometric considerations. *ETS Research Report Series*, 2013(2), i-36. <https://doi.org/10.1002/j.2333-8504.2013.tb02348.x>
- von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: Lessons learned from developing a collaborative science assessment prototype. *Computers in Human Behavior*, 76, 631–640. <https://dx.doi.org/10.1016/j.chb.2017.04.059>
- Voogt, J., Erstad, O., Dede, C., & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, 29(5), 403–413. doi:10.1111/jcal.12029
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467. <https://dx.doi.org/10.1007/s11412-017-9267-5>
- 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(2), 323–343. <https://dx.doi.org/10.1007/s11251-012-9230-9>
- Yeh, Y. C. (2010). Analyzing online behaviors, roles, and learning communities via online discussions. *Journal of Educational Technology & Society*, 13(1), 140–151. Retrieved from <http://www.jstor.org/stable/jeductechsoci.13.1.140>
- Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., & Katz, I. R. (2014). Assessing science inquiry skills using dialogues. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014)*, 5–9 June 2014, Honolulu, HI, USA (pp. 625–626). Springer. [https://dx.doi.org/10.1007/978-3-319-07221-0\\_84](https://dx.doi.org/10.1007/978-3-319-07221-0_84)

## Appendix A

The analyses proceeded with 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 1 for the five-cluster model. The ANOVA revealed significant differences among clusters for all of the six GCA variables at the  $p < .0001$  level for the five-cluster model. Tukey’s HSD *post hoc* comparisons for the five-cluster model are reported in Table 2, where we can see that the observed differences in GCA profiles across the clusters were, for the majority, significantly distinct.

**Table 1. Five-cluster Model Means, Standard Deviations, and ANOVA for the Six GCA Variables**

GCA Measures	Cluster 1: Driver <i>n</i> = 168	Cluster 2: Follower <i>n</i> = 181	Cluster 3: Influential <i>n</i> = 173	Cluster 4: Lurker <i>n</i> = 242	Cluster 5: Detached <i>n</i> = 203	<i>F</i> -value
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	
Participation	0.7 (0.63)	-0.9 (0.59)	0.28 (0.67)	-0.8 (0.66)	0.93 (0.64)	353.50***
Social Impact	0.74 (0.98)	0.8 (0.82)	0.33 (0.7)	-0.7 (0.56)	-0.77 (0.55)	238.80***
Overall Responsivity	0.32 (1.03)	0.78 (0.91)	0.11 (0.72)	-0.35 (0.99)	-0.64 (0.6)	221.40***
Internal Cohesion	0.91 (1)	0.58 (0.82)	0.42 (0.64)	-0.75 (0.54)	-0.74 (0.54)	80.75***
Newness	0.7 (1.05)	-0.75 (0.61)	0.17 (0.48)	-0.56 (0.69)	0.61 (1.04)	131.40***
Communication Density	0.87 (0.9)	-0.57 (0.72)	-0.59 (0.56)	-0.23 (0.8)	0.57 (1.01)	121.80***

Note: ANOVA *df* = 4,962; \*\*\*  $p < .0001$

**Table 2. Tukey-HSD P-Values for the Pairwise Comparisons for the GCA Measures Across the Five-Cluster Solution**

Cluster Comparison	GCA Variables					
	Participation	Social Impact	Overall Responsivity	Internal Cohesion	Newness	Communication Density
2 vs. 1	$p = .001$	$p < .001$	$p = .94$	$p < .001$	$p < .001$	$p < .001$
3 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p = .17$	$p < .001$	$p < .001$
4 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
5 vs. 1	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = .82$	$p < .01$
3 vs. 2	$p < .001$	$p = .20$	$p < .001$	$p < .001$	$p < .001$	$p = .99$
4 vs. 2	$p = .49$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
5 vs. 2	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
4 vs. 3	$p = .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
5 vs. 3	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
5 vs. 4	$p < .001$	$p < .999$	$p = .81$	$p < .05$	$p < .001$	$p < .001$