# From Discourse to Dynamics: Understanding Team Interactions Through Temporally Sensitive NLP

Seehee Park*
University of California, Irvine
seeheep@uci.edu

Danielle Shariff*
University of California, Irvine
dsharif1@uci.edu

Mohammad Amin Samadi
University of California, Irvine
masamadi@uci.edu

Nia Nixon
University of California, Irvine
dowelln@uci.edu

Sidney D'Mello
University of Colorado Boulder
sidney.dmello@colorado.edu

## ABSTRACT

Collaborative Problem Solving (CPS) is a vital 21st-century skill that integrates social and cognitive processes to achieve shared goals. Despite its importance, understanding how communication dynamics shape individual learning outcomes in CPS tasks remains a challenge, particularly in virtual settings. To address this gap, this study analyzes discourse patterns using Group Communication Analysis (GCA), a temporally sensitive computational methodology for quantifying team interactions. Our work takes a novel approach by identifying discourse patterns associated with student performance. We examined the interaction patterns of 279 undergraduate students as they engaged in a virtual physics learning game, using GCA measures such as participation, internal cohesion, responsivity, social impact, newness, and communication density. Our findings suggest that students who introduced more novel contributions (i.e., newness), established common ground (i.e., overall responsivity) and engaged in denser communication showed significantly greater improvements in their post-test scores. These results provide valuable insights into optimizing virtual collaboration in educational settings, highlighting the role of novel information exchange between learners. This study emphasizes the importance of examining individual contributions and group dynamics, and proposes future research directions to the EDM community.

## Keywords

Collaborative Problem Solving, Learning Analytics, Group Communication Analysis, Discourse Analytics, Virtual Learning Environment

## 1. INTRODUCTION & BACKGROUND

*Both authors contributed equally to this work.

Collaborative problem solving (CPS) is defined as two or more individuals engaging in a coordinated effort to construct and maintain a joint solution to a problem [31]. Recognized as a critical 21st-century competency [3], CPS plays a vital role in education by fostering teamwork and enhancing learning outcomes [17, 26]. Effective CPS requires team members to collaboratively define problems, set goals, and monitor their progress [17], while also engaging in key processes such as task distribution, communication, information sharing, and consensus-building. Given its complexity, successful CPS relies on the integration of both social and cognitive processes [13]. These processes are deeply intertwined with social processes (e.g., responding to peers, asking clarifying questions) and cognitive processes (e.g., proposing ideas, engaging in task discussions) [42]. Frameworks such as PISA and ATC21S emphasize the interplay between social and cognitive dimensions of collaboration, both of which are essential for successful CPS outcomes [39, 42]. Social aspects involve the interpersonal processes that facilitate collaboration, such as establishing shared understanding, negotiating meaning. On the other hand, cognitive aspects involve problem-solving processes such as planning, reasoning, generating hypotheses, monitoring progress, and integrating knowledge. These two dimensions are deeply intertwined in collaborative problem solving (CPS), requiring team members to communicate effectively, negotiate ideas, resolve conflicts, and co-construct strategies to achieve shared goals [1]. Gaining a deeper understanding of how social and cognitive processes interact during team communication is essential for advancing research and practice in CPS.

Communication is central to collaboration, making language critical for evaluating learning processes. In educational research, discourse analysis has long been used to examine learners' social, cognitive, and affective states [4, 12, 15]. Prior research has shown that linguistic coordination, cohesion, and discourse dynamics serve as key indicators of socio-cognitive processes in group interactions [10, 27, 34]. In the context of collaborative learning, analyzing team communication provides valuable insight into both individual collaboration skills [1] and group-level processes such as information sharing, coordination, negotiation, and progress monitoring [12]. As the volume of discourse data continues to grow, computational linguistic approaches have been used to identify communication patterns and team outcomes [17]. With ongoing efforts, educational data mining (EDM) has offered

valuable insights into learner interactions and performance, creating new opportunities to systematically examine CPS processes across diverse educational settings [37, 35, 32, 8]. Building on this, we aim to explore how communication dynamics are associated with learning in collaborative settings, as understanding these relationships can provide deeper insight into the mechanisms that drive effective team learning.

Advancements in natural language processing (NLP) have enhanced the ability to analyze large volumes of textual data, providing valuable insights into collaborative discourse [21]. However, despite these advancements, there remains a need for more fine-grained temporal discourse analysis to better understand conversational dynamics in teams [11]. While NLP techniques have been widely applied to assess discourse quality at scale, many approaches focus on static or aggregate measures, limiting their ability to capture the evolving nature of discourse throughout collaboration [11, 30]. In the context of collaborative problem solving (CPS), capturing these evolving discourse patterns are particularly important, as effective collaboration emerges through continuous interaction among team members. To address this, we employ Group Communication Analysis (GCA), a computational linguistic method that analyzes discourse as a temporally structured sequence of contributions. GCA enables individuals to have dependencies in communication, offering a nuanced understanding of how collaboration unfolds in multi-party interactions [11].

In the context of collaborative problem solving (CPS), studies have been trying to understand social-cognitive dimensions such as teamwork, negotiation, and coordination by analyzing surface-level lexical features (e.g., word frequency) and conversational dynamics (e.g., turn-taking patterns) [29]. NLP-based approaches link these observable features to underlying cognitive processes like problem exploration or solution generation, and further relate them to CPS outcomes such as task performance [29, 44]. While Social Network Analysis (SNA) offers insight into structural relations among participants, it also shows limitations in capturing deeper interpersonal and socio-cognitive dynamics embedded in discourse interactions [11]. Although recent work has examined temporal dynamics in team communication, there remains a significant need for more nuanced, temporally sensitive approaches that can reveal how these social and cognitive processes unfold and interact over time. To address these limitations, Group Communication Analysis (GCA) builds on prior NLP and network approaches by emphasizing the temporal dynamics of multi-party interactions. GCA combines sequential alignment and coordination measures with semantic models (e.g., Word2Vec, GloVe) to trace how meaning and coordination evolve throughout discourse [11, 29]. By capturing fluctuations in socio-cognitive processes over time, GCA provides a fine-grained lens for examining how learners sustain engagement, negotiate meaning, and build shared knowledge during CPS interactions (for more details go to section 2.5).

Drawing on this, the current study explores learning gains in relation to conversational dynamics measured through GCA. Specifically, we applied six GCA measures i.e., Newness, Communication Density, Social Impact, Internal Cohesion, Participation, and Overall Responsivty to examine how communication patterns shape teamwork and learning outcomes. To achieve this, we determined individual learning gains by comparing pre- and post-test performance, and then categorized participants into low, medium, and high gain groups using quantile estimations. We analyze both interpersonal and intrapersonal conversational data to uncover how group communication patterns correlate with learning gains and how individuals in different gain levels behave and communicate differently in teams. In doing so, our work sheds light on the interplay between team processes and performance that distinguish high-performing individuals and foster their success and growth. Building on these insights, we address the following research question:

1. Are there significant differences in conversational dynamics among participants with low, medium, and high learning gains, as determined by score gain percentiles?

## 2. METHODS
### 2.1 Participants
The study involved 279 undergraduate students from two US public universities. Participants were 56% female, averaging 21.73 years old from University 1. Participants self-reported the following race/ethnicity: 50% White, 26% Hispanic or Latino, 18% Asian, 3% Black or African American, 1% Native American or Alaska Native, and 2% "Other" and among them, 75% were native English speakers. Participants were assigned to triads with thirty headset microphones, partitioned in different corners of the same room or located in different rooms, depending on the university where the data were collected. All collaboration occurred via Zoom video-conferencing software. Zoom recordings of all collaborations were retained for analysis.

### 2.2 Task
Students engaged in a collaborative problem-solving task using the educational game Physics Playground [36], which teaches basic physics concepts such as Newton's laws, energy transfer, and torque. During game-play, teams solved levels over three 15-minute sessions (warm-up, block 1 and block 2), totaling 45 minutes of game-play. In Physics Playground, students complete levels by using a mouse to draw simple devices, such as ramps, levers, pendulums, and springboards, to solve physics-based challenges. Within each triad, team members rotated roles: one student acted as the Controller, sharing their screen via Zoom, while the other two students served as Contributors, communicating freely through headsets. Spoken communication was encouraged, and the use of the chat box was restricted. This study specifically analyzed two 15-minute experimental blocks, excluding the initial warm-up session [42, 41].

### 2.3 Data Collection
This study utilizes text data from pre- and post- tests, including group communication transcripts derived from audio recordings, to analyze interaction patterns. The pre- and post-test scores were collected before and after engaging in the CPS task. Participants completed a ten-item physics pre-test, developed by experts, to assess their understanding of energy transfer and torque properties, aligning

Table 1: Definitions of GCA measures

| Measure | Description |
|---------|-------------|
| Participation | Mean participation of an individual relative to the average of the group of its size. |
| Social Impact | Measure of how contributions initiated by the corresponding paritcipant have triggered follow-up responses. |
| Overall Responsivity | Measure of the tendency of an individual to respond, or not, to the contributions of their peers in the group. |
| Internal Cohesion | Measure of how consistent an individual is with their own recent contributions. |
| Newness | Measure of how likely for an individual to provide new information or to echo existing information. |
| Communication Density | Measure of the amount of semantically meaningful information in utterances. |

with the Physics Playground levels used in the CPS activity. This pre-test established a baseline measure of their physics knowledge. Following the CPS task, participants completed a post-test, which was a parallel version of the pretest [40].

## 2.4 Score Gain

Initially, score gains were measured as the raw difference between post-test and pre-test scores. However, this approach had limitations, particularly for participants with perfect pre-task scores (showing minimal or negative gains) and for those with significant post-task score drops, potentially skewing the results. To address these issues, we adopted a weighted gain score metric [46]:

$$G_4 = (post - pre) \times \left( \frac{pre}{\mu} \right) \qquad (1)$$

In this formula, $\mu$ represents the expected mean learning gain, set to 50 (50%) to align with common benchmarks for intervention effectiveness and to balance initial performance disparities while minimizing extreme score changes [46].

## 2.5 Group Communication Analysis

This study will use a temporally sensitive NLP approach, Group Communication Analysis (GCA), which provides six primary measures of team interaction: Participation, Internal Cohesion, Responsivity, Social Impact, Newness, and Communication Density (see Table 1). These metrics quantify the dynamics of individual contributions within collaborative problem-solving contexts over time. GCA combines artificial intelligence methods, such as computational semantic models of cohesion, with temporally sensitive semantic analyses inspired by the cross- and auto-correlation measures from time-series analysis [13, 11, 14]. These semantic space models, which rely on advanced artificial intelligence techniques, may be constructed via Latent Semantic Analysis (LSA) [23], a classic matrix-factorization method, or more current artificial neural network word embedding models such as Skip-gram (i.e. Word2vec, [25]) or Global Vectors of Words (i.e., "GloVe", [28]). Using this approach, GCA allows researchers to quantify discourse as a dynamic and evolving socio-cognitive process that lies in the interaction between learners' communicative contributions.

Specifically, Group Communication Analysis (GCA) provides a comprehensive framework for understanding effective group interactions [11], capturing individual and group dynamics that shape collaboration. Participation reflects engagement and willingness to contribute, a fundamental prerequisite for teamwork [6, 19], while internal cohesion assesses the semantic consistency of contributions, balancing integration with adaptability [2]. Responsivity and social impact reveal how individuals align their inputs with peers and influence group dialogue, fostering shared understanding and sustained collaboration [43, 45]. Newness highlights the introduction of novel ideas, critical for innovation and avoiding stagnation [7], while communication density measures the efficiency of meaningful exchanges, crucial for high-performing teams [16]. For example, high newness scores might reflect speakers who introduce new words or concepts that have not been discussed previously, or ideas that expand on existing information, while low newness scores might be indicative of speakers who echo knowledge already present in the discourse. High communication density scores, on the other hand, signal concise (i.e., fewer words) yet information-rich contributions; a low communication density score might reflect contributions that are verbose and contain less semantically meaningful information per word. In other words, the speaker may use more words to convey relatively little new or substantive content. These measures offers nuanced insights into the interplay of individual contributions and collective outcomes. Operational definitions of the GCA measures are provided in Table 1.

## 3. STATISTICAL ANALYSIS

To address our research question, we first stratified the data into three groups based on participants' score gains percentile. The three subgroups were defined based on values at the $33rd$, $67th$, and $100th$ percentiles using the Hyndman and Fan quantile estimation method [20]. Of the original sample ($n = 279$), 16 participants were excluded due to missing data. The final sample ($n = 263$) included 146 participants with low score gains, 31 with medium, and 86 with high gains. This uneven distribution reflects how Hyndman and Fan's method prioritizes accuracy in estimating the data distribution over enforcing equal group sizes. Descriptive statistics for the measurements across the three groups

Table 2: Descriptive Statistics for low, med, and high in score gains

| Variable | Low ($n = 146$) Mean ($SD$) | Med ($n = 31$) Mean ($SD$) | High ($n = 86$) Mean ($SD$) | All ($N = 263$) Mean ($SD$) |
|---|---|---|---|---|
| PhysicsScorePre | 7.09 (1.70) | 4.55 (1.40) | 6.35 (1.65) | 6.55 (1.83) |
| PhysicsScorePost | 6.25 (1.91) | 5.81 (0.87) | 8.44 (1.24) | 6.91 (1.94) |
| Score Gain (Adjusted) | -0.12 (0.17) | 0.09 (0.03) | 0.24 (0.08) | 0.02 (0.22) |
| Participation | 0.02 (0.07) | 0.01 (0.07) | 0.03 (0.09) | 0.02 (0.08) |
| Internal Cohesion | 0.19 (0.04) | 0.18 (0.04) | 0.18 (0.04) | 0.19 (0.04) |
| Overall Responsivity | 0.15 (0.03) | 0.17 (0.03) | 0.16 (0.03) | 0.16 (0.03) |
| Social Impact | 0.15 (0.03) | 0.16 (0.02) | 0.15 (0.02) | 0.15 (0.02) |
| Newness | 0.71 (2.16) | 2.27 (10.36) | 4.33 (15.43) | 2.08 (9.74) |
| Comm Density | 0.47 (1.58) | 1.69 (7.88) | 2.67 (9.40) | 1.33 (6.18) |

*Note.* The table presents the mean and standard deviation (Mean $\pm SD$) for variables across three Gain groups: High, Medium, and Low. Comm Density refers to communication density. Values are rounded to two decimal places.

Table 3: ANOVA Results Summary with GCA measures and adjusted score gains

| Variable | df | SS | MS | $F$-value | $p$ |
|---|---|---|---|---|---|
| Newness | 2 | 709.00 | 354.70 | 3.82 | **0.02**\* |
| Comm Density | 2 | 268.00 | 133.81 | 3.57 | **0.03**\* |
| Social Impact | 2 | 0.00 | 0.00 | 0.18 | 0.83 |
| Internal Cohesion | 2 | 0.00 | 0.00 | 0.13 | 0.88 |
| Participation | 2 | 0.01 | 0.00 | 0.99 | 0.37 |
| Overall Responsivity | 2 | 0.00 | 0.00 | 3.13 | **0.04**\* |

*Note.* Signif. codes: df= degrees of freedom, SS= Sum of Squares, MS= Mean Square, $p$ = significance value

are presented in Table 2. Second, we performed one-way Analysis of Variance (ANOVA) tests to determine whether significant differences existed in GCA measures across the three learning gain groups. Using Tukey-HSD comparison, post-hoc analyses were conducted to identify specific differences between the *Low*, *Medium*, and *High Gains* groups.

## 4. RESULTS

As described in Section 3, we conducted a series of one-way analyses of variance (ANOVAs) to examine whether the six GCA variables differed across the *Score Gain* subgroups: *Low*, *Medium*, and *High*. The results, summarized in Table 3, indicated statistically significant differences for three of the six GCA variables. There was a significant effect of score gain subgroup on newness, $F(2, 260) = 3.82, p = .02$. A significant effect was also found for communication density, $F(2, 260) = 3.57, p = .03$. Finally, overall responsivity differed significantly across subgroups, $F(2, 260) = 3.13, p = .04$. No significant differences were found for social impact, internal cohesion, or participation.

Post hoc comparisons using Tukey's HSD test revealed specific group differences (see Table 4). For newness, the *High Gains* group ($M = 4.33$, $SD = 15.43$) scored significantly higher than the *Low Gains* group ($M = 0.71$, $SD = 2.16$), $p = .02$. For communication density, the *High Gains* group ($M = 2.67$, $SD = 9.40$) again scored significantly higher than the *Low Gains* group ($M = 0.47$, $SD = 1.58$), $p = .02$. Lastly, for overall responsivity, the *Medium Gains* group ($M = 0.17$, $SD = 0.03$) showed significantly higher scores than the *Low Gains* group ($M = 0.15$, $SD = 0.03$), $p = .03$.

Effect sizes were estimated using Cohen's d for each pairwise comparison (Table 4). The Low–High comparison for newness showed a small-to-medium effect size, $d = 0.38$, 95% CI [0.11, 0.64]. For communication density, the same comparison yielded $d = 0.38$, and the Medium–Low comparison for overall responsivity showed the largest effect, $d = -0.51$, 95% CI [–0.98, –0.05].

## 5. DISCUSSION

In this study, we explored temporal communication dynamics and their association with learning gains. We used pre- and post-test scores to calculate learning gains and investigated their relationship with GCA variables. Although several comparisons yielded statistically significant results, the effect sizes ranged from small to medium. This implies that the relationships are meaningful but may have limited practical significance. This could be attributed to our smaller sample size that limits the extent to which we can generalize beyond the study's sample. In this context, our findings suggest that students who introduced more novel contributions and engaged in more semantically meaningful communication showed significantly greater improvements in their post-task test scores. Moreover, frequently sharing new ideas during the problem-solving task was associated with higher gains in performance. We found that *Low* and *High* performing individuals were less responsive to peer contributions than the *Medium Gains* group, showcasing less overall responsivity. Participants in the *Low Gains* group tended to contribute less new information, while those in the *Medium Gains* group generally responded to their peers but did not always add new ideas. The *High Gains* group, though responding less frequently to peer information, often provided novel and meaningful contributions when they did engage. Lastly, all three performance groups had similar lev-
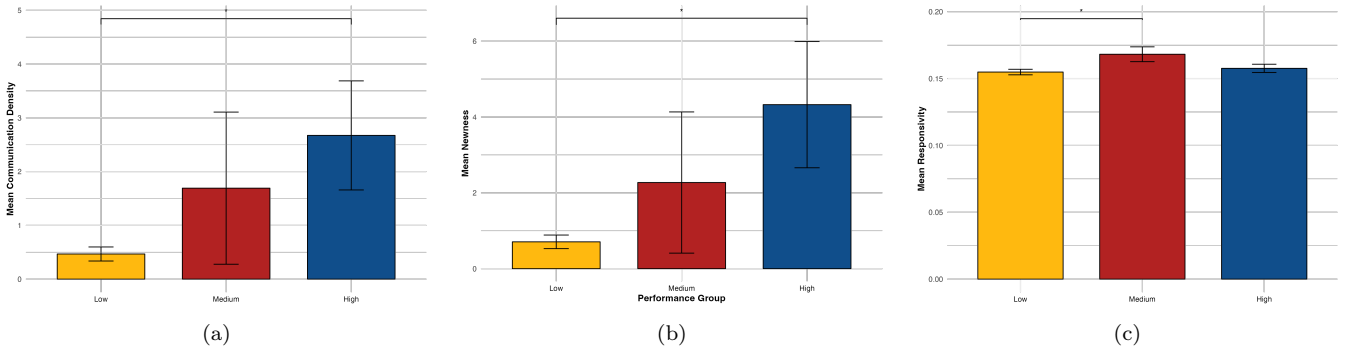
(a)    (b)    (c)

Figure 1: Mean comparisons and confidence intervals of GCA measures across different gain groups. The plots illustrate each GCA measure included in the post-hoc analysis: (a) Communication Density, (b) Newness, and (c) Responsivity. An asterisk (*) denotes significant differences between groups in the post-hoc analysis. Note: The scaling of each plot varies, and values are not directly comparable across measures.

Table 4: Tukey HSD Post-Hoc Results with Effect Sizes

| Variable | Comparison | MD | SE | 95% CI | p | d |
|---|---|---|---|---|---|---|
| **Newness** | Low - High | -3.62 | 1.58 | [-6.70, -0.53] | **0.02*** | 0.38 |
| | Med - High | -2.05 | 2.43 | [-6.81, 2.71] | 0.57 | 0.14 |
| | Med - Low | 1.56 | 2.29 | [-2.93, 6.06] | 0.69 | -0.33 |
| **Comm Density** | Low - High | -2.20 | 1.00 | [-4.17, -0.24] | **0.02*** | 0.38 |
| | Med - High | -0.98 | 1.54 | [-4.00, 2.04] | 0.72 | 0.11 |
| | Med - Low | 1.22 | 1.46 | [-1.63, 4.08] | 0.57 | -0.34 |
| **Overall Responsivity** | Low - High | -0.00 | 0.00 | [-0.01, 0.00] | 0.73 | 0.11 |
| | Med - High | 0.01 | 0.00 | [-0.00, 0.02] | 0.15 | -0.36 |
| | Med - Low | 0.013 | 0.00 | [0.00, 0.02] | **0.03*** | -0.51 |

*Note.* Significant results marked with * ($p < .05$). $MD$ = Mean difference; $SE$ = Standard error; $d$ = Cohen's $d$, using pooled SD. CI = confidence interval, $p$ = significance value

els of participation. Participation was not significantly associated with learning gains; this finding further highlights the agency of more nuanced metrics such as newness, responsivity, and communication density, in comparison to characteristically static measures like participation or number of messages.

Research in the team science literature repeatedly demonstrates that typical measures of participation are oversimplistic and do not tell us how team dynamics change and develop over time [9]. Instead of quantifying the number or frequency of an individual's utterances, effective engagement is better measured by the substance and impact of their discourse. That is, the productivity and value of their contributions matter more than how often they speak [10]. Our findings are also consistent with previous research indicating that team members benefit from generating, sharing, and integrating new information that constructively contributes to the dialogue or task at hand [7]. Effective knowledge sharing, particularly from those who not only contribute novel information but also introduce practically valuable ideas, significantly shapes successful task completion [47].

In the context of our study, we qualitatively explored why newness and communication density were particularly relevant. Upon reviewing transcripts, we found that participants with high newness and communication density scores were rarely among the role of controllers. They often directed the controller and made significant contributions to

the team. These participants tended to steer the discussion productively, driving the dynamics of the group interaction towards task completion. For instance, a participant with both high newness and communication density instructed their teammate, who was the controller:

> "If you try drawing it straight down instead...instead of like off to the side that way it won't swing and then maybe we can time the swing of the ball down there."

Here, the participant offers an alternative solution for the physics game level. In a separate instance, another student guides the Controller by providing sequential instructions:

> "Yeah, so just draw a heavy object again, just like you did and keep the, um ball on the right and then let the object fall and then when the lever's all the way down delete the object."

These utterances, and many similar ones, illustrate the importance of holistically understanding virtual teams dynamics to effectively assess learning gains, gaps, and individual needs. With one participant controlling the screen, the distribution of roles does not mirror the more organic flow of a face-to-face conversation. Typically, collaborative groups

414

are formed without predefined, rigid roles assigned [22]. Participants then take on flexible and emergent roles during the CPS process, which reflects the dynamical and complex nature of CPS spaces [24]. Team interactions are impacted by participants' emotions, group conflict, task challenges and task progression, rendering it difficult to confine a participant to a single or specific role. In our study, defined roles likely influenced how participants interacted with each other. For example, a predefined role, like the controller, prompted other participants to direct the actions of the controller, in which they more actively contributed to provide support and make recommendations.

Our findings make it evident that successful learning in teams is shaped by evolving and complex dynamics, of which discourse serves as a driving force for collaboration. As technological advancements and broader societal shifts continue to redefine CPS environments, there is an increasing need for theoretical and methodological approaches that incorporate contextual features and dynamic interactions. In our study, structured elements such as 15-minute time blocks, rotating roles, and multi-modal communication created a unique system of interactions, requiring participants to navigate platform constraints and rapidly evolving tasks. Advancing our ability to study team interactions not only provides us with a data-driven understanding of how individuals engage online, but also informs the design of more meaningful, supportive and successful team learning [18, 36, 38]. In doing so, this work contributes to ongoing efforts within the EDM community to build more contextually grounded and temporally responsive models of collaborative learning.

While the current study has significant implications for observing successful problem-solving teamwork in the context of virtual environments, there are some limitations that should be considered. Extreme variability within similarly structured subgroups made detecting a clear and reliable pattern difficult. Even for participants who actively engaged with demonstrating high communication density, introducing novel contributions, and incorporating peer information, learning gains may have appeared low or negative because of an initially high pre-task score. This has made it increasingly challenging to predict communication patterns that endure and remain consistent. Additionally, we emphasize that temporally and contextually sensitive approaches are crucial for capturing how information is introduced and leveraged in short but information-dense exchanges. Current NLP techniques often lack this nuance, highlighting the need for more refined analytic strategies—particularly those capable of assessing team and individual success dynamically, rather than relying solely on static test scores.

Building on these findings, we consider their implications for future interventions and real-world applications. GCA measures of newness, overall responsivity, and communication density offer direction for optimizing CPS through automated tools and human-facilitated interventions. For instance, automated feedback systems could leverage GCA metrics in real-time to monitor group interactions and provide tailored prompts to students, nudging them toward more productive communication behaviors. A chatbot or virtual facilitator [5, 33] could be designed to detect moments of low participation, lack of idea novelty, or low se-

mantic density and intervene with questions or suggestions to stimulate deeper engagement (e.g., "Can someone propose a new approach?" or "How does this idea build on what was said earlier?"). Future research should explore the design and evaluation of such interventions, including testing how different forms of real-time feedback or instructor support influence communication dynamics and learning outcomes. By integrating GCA metrics into automated tools or instructional practices, educators can move beyond static participation measures and foster more meaningful collaboration. In conclusion, this study enhances our understanding of how communication dynamics influence learning in collaborative settings, emphasizing their role in shaping learning outcomes. These findings inform the development of educational interventions aimed at supporting and facilitating deeper learning in CPS contexts.

# 6. ACKNOWLEDGMENT

# 7. REFERENCES

[1] J. Andrews-Todd and C. M. Forsyth. Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in human behavior*, 104:105759, 2020.

[2] B. Barron. Achieving coordination in collaborative problem-solving groups. *The journal of the learning sciences*, 9(4):403–436, 2000.

[3] J. Burrus, T. Jackson, N. Xi, and J. Steinberg. Identifying the most important 21st century workforce competencies: An analysis of the occupational information network (o*net). *ETS Research Report Series*, 2013(2), 2013. Pages i–55.

[4] W. Cade, N. Dowell, A. Graesser, Y. Tausczik, and J. Pennebaker. Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, pages 399–400, London, UK, 2014. International Educational Data Mining Society.

[5] Z. Cai, S. Park, N. Nixon, and S. Doroudi. Advancing knowledge together: Integrating large language model-based conversational ai in small group collaborative learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024.

[6] E. Care, C. Scoular, and P. Griffin. Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, 29(4):250–264, 2016.

[7] M. T. Chi and M. Menekse. Dialogue patterns in peer collaboration that promote learning. *Socializing intelligence through academic talk and dialogue*, 1(2):263–274, 2015.

[8] H. Chopra, Y. Lin, M. A. Samadi, J. G. Cavazos, R. Yu, S. Jaquay, and N. Nixon. Semantic topic chains for modeling temporality of themes in online student discussion forums. *International Educational Data*

*Mining Society*, 2023.

[9] F. Delice, M. Rousseau, and J. Feitosa. Advancing teams research: What, when, and how to measure team dynamics over time. *Frontiers in psychology*, 10:1324, 2019.

[10] N. Dowell, T. Nixon, and A. Graesser. Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multi-party interactions. *Behavior Research Methods*, 51, 01 2018.

[11] N. Dowell, O. Poquet, and C. Brooks. Applying group communication analysis to educational discourse interactions at scale. In *Proceedings of the 13th International Conference of the Learning Sciences (ICLS 2018)*, pages 579–586, London, UK, 2018. International Society of the Learning Sciences (ISLS).

[12] N. M. Dowell, A. C. Graesser, and Z. Cai. Language and discourse analysis with coh-metrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, 3(3):72–95, 2016.

[13] N. M. Dowell, Y. Lin, A. Godfrey, and C. Brooks. Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes: A group communication analysis. *Journal of Learning Analytics*, 7(1):38–57, 2020.

[14] N. M. Dowell and O. Poquet. Scip: Combining group communication and interpersonal positioning to identify emergent roles in scaled digital environments. *Computers in Human Behavior*, 119:106709, 2021.

[15] S. K. D'Mello, N. Dowell, and A. Graesser. Unimodal and multimodal human perceptionof naturalistic non-basic affective statesduring human-computer interactions. *IEEE Transactions on Affective Computing*, 4(4):452–465, 2013.

[16] J. C. Gorman, P. W. Foltz, P. A. Kiekel, M. J. Martin, and N. J. Cooke. Evaluation of latent semantic analysis-based measures of team communications content. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 424–428, Los Angeles, CA, 2003. Sage Publications.

[17] A. C. Graesser, N. Dowell, A. J. Hampton, A. M. Lippert, H. Li, and D. W. Shaffer. Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 program for international student assessment. In B. Harrison and C. D. Schunn, editors, *Building Intelligent Tutoring Systems for Teams*, volume 19, pages 173–211. Emerald Publishing Limited, Bingley, UK, 2018.

[18] Z. T. Harvey, G. A. Loomis, S. Mitsch, I. C. Murphy, S. C. Griffin, B. K. Potter, and P. Pasquina. Advanced rehabilitation techniques for the multi-limb amputee. *Journal of Surgical Orthopaedic Advances*, 21(1):50–57, 2012.

[19] S. Hrastinski. What is online learner participation? a literature review. *Computers & education*, 51(4):1755–1765, 2008.

[20] R. J. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.

[21] S. Joksimović, O. Poquet, V. Kovanović, N. Dowell, C. Mills, D. Gašević, S. Dawson, A. C. Graesser, and C. Brooks. How do we model learning at scale? a systematic review of research on moocs. *Review of Educational Research*, 88(1):43–86, 2018.

[22] P. A. Kirschner and G. Erkens. Toward a framework for cscl research. *Educational Psychologist*, 48(1):1–8, 2013.

[23] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch. *Handbook of latent semantic analysis*. Psychology Press, 2007.

[24] Z. Mao, X. Li, and Y. Li. Identifying emergent roles and their relationship with learning outcomes and collaborative problem-solving skills. *Thinking Skills and Creativity*, 54:101642, 2024.

[25] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26:3111–3119, 2013.

[26] OECD. Pisa 2015 collaborative problem solving framework, 2017.

[27] S. Park, N. Nixon, S. D'Mello, D. Shariff, and J. Choi. Understanding collaborative learning processes and outcomes through student discourse dynamics. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference*, pages 938–943, 2025.

[28] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[29] S. Pugh, S. K. Subburaj, A. R. Rao, A. E. Stewart, J. Andrews-Todd, and S. K. D'Mello. Say what? automatic modeling of collaborative problem solving skills from student speech in the wild. In *Proceedings of the 14th international conference on educational data mining*, pages 256–266. Educational Data Mining, 2021.

[30] M. Robeer, F. Bex, A. Feelders, and H. Prakken. Explaining model behavior with global causal analysis. In *World Conference on Explainable Artificial Intelligence*, pages 299–323. Springer, 2023.

[31] J. Roschelle. The construction of shared knowledge in collaborative problem solving. *Computer Supported Collaborative Learning/Springer-Verlag*, 1995.

[32] M. A. Samadi, J. G. Cavazos, Y. Lin, and N. Nixon. Exploring cultural diversity and collaborative team communication through a dynamical systems lens. *International Educational Data Mining Society*, 2022.

[33] M. A. Samadi, S. JaQuay, J. Gu, and N. Nixon. The ai collaborator: Bridging human-ai interaction in educational and professional settings. *arXiv preprint arXiv:2405.10460*, 2024.

[34] M. A. Samadi, S. Jaquay, Y. Lin, E. Tajik, S. Park, and N. Nixon. Minds and machines unite: deciphering social and cognitive dynamics in collaborative problem solving with ai. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pages 885–891, 2024.

[35] M. A. Samadi and N. Nixon. Cultural diversity in team conversations: A deep dive into its effects on cohesion and team performance. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 821–827, 2024.

[36] V. J. Shute, M. Ventura, and Y. J. Kim. Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research*, 106(6):423–430, 2013.

[37] G. Siemens and R. S. d. Baker. Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 252–254, 2012.

[38] T. D. Snyder, C. de Brey, and S. A. Dillow. Digest of education statistics 2017. Technical Report NCES 2018-070, National Center for Education Statistics, 2019.

[39] M. Stadler, K. Herborn, M. Mustafić, and S. Greiff. The assessment of collaborative problem solving in pisa 2015: An investigation of the validity of the pisa 2015 cps tasks. *Computers & Education*, 157, 2020. Article 103964.

[40] A. E. Stewart, M. J. Amon, N. D. Duran, and S. K. D'Mello. Beyond team makeup: Diversity in teams predicts valued outcomes in computer-mediated collaborations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[41] C. Sun, V. J. Shute, A. Stewart, J. Yonehiro, N. Duran, and S. D'Mello. Towards a generalized competency model of collaborative problem solving. *Computers & Education*, 143:103672, 2020.

[42] C. Sun, V. J. Shute, A. E. Stewart, Q. Beck-White, C. R. Reinhardt, G. Zhou, N. Duran, and S. K. D'Mello. The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, 128:107120, 2022.

[43] D. Suthers. From contingencies to network-level phenomena: Multilevel analysis of activity and actors in heterogeneous networked learning environments. In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 368–377, 2015.

[44] I. Vazquez, R. Gonzalez, and P. A. Gloor. Measuring collaboration through nlp-assessed personality characteristics of team members. `https://ssrn.com/abstract=5181190`, 2025. Available at SSRN: `https://ssrn.com/abstract=5181190`.

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

[46] S. Westphale, J. Backhaus, and S. Koenig. Quantifying teaching quality in medical education: The impact of learning gain calculation. *Medical Education*, 56(3):312–320, 2022.

[47] G. M. Wittenbaum, A. B. Hollingshead, and I. C. Botero. From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs*, 71(3):286–310, 2004.