Proceedings of the Human Factors and Ergonomics Society Annual Meeting
2023, Vol. 67(1) 290–295

DOI: 10.1177/21695067231192643
journals.sagepub.com/home/pro
Sage

Toward the Validation of a Novel Measure of Individuals' Influence During Team Collaborations

Julie L. Harrison[1], Jamie C. Gorman[2], Jason G. Reitman[3], and Sidney D'Mello[3]

Abstract

Collaborative problem solving (CPS) is an important skill in the modern workforce, and due to its interactive nature, is challenging to assess. The present study builds on work in team sciences to provide initial validation for a metric that quantifies *CPS influence*—the extent to which each individual contributes toward the team's CPS processes—using average mutual information (AMI). The measure is investigated in teams collaborating in a computer programming task, where one teammate was assigned to a controller role (i.e., the only person who engaged with the task interface directly). Results suggest the controller had more influence over the team's CPS processes than the other participants in the triad, providing initial validation for the influence metric. Future work will investigate the measure in classrooms and multiple modalities, and extend the metric in real-time to understand how influence fluctuates over the course of collaboration.

Keywords

collaboration, team science, team performance, classroom education

Introduction

Professional settings now increasingly require employees to work collaboratively to achieve institutional goals (Graesser et al., 2018). Accordingly, collaborative problem solving (CPS) has been identified by the Organisation for Economic Co-operation and Development (OECD) as a skill implicated in the success of the global economy. CPS involves two or more agents working interdependently to solve a novel problem, by arriving at a set of steps to transform a given state into a goal state (Sun et al., 2020). Given its interactive nature, CPS is a challenging construct to assess at the level of the individual, making it difficult to quantify using traditional methods in educational assessment. While numerous measurement frameworks have been proposed that identify CPS skills from task artifacts, such as chat logs or transcripts (e.g., OECD, 2017; Sun et al., 2020), CPS quantification remains elusive due to the complexity of naturalistic social interactions. The present study utilizes one such measurement framework—The Generalized Competency Model of CPS (Sun et al., 2020)—to develop a novel metric that quantifies the extent to which an individual influences the group's output. Future iterations of this proposed metric allow for the near-real-time assessment of teammates' influence on collaborative team processes during learning and training.

An important assumption behind our approach is that the function and behavior of a team must be considered at both the individual and team levels. For example, *social loafing*—the phenomenon wherein a teammate(s) shifts the burden of the work and responsibility to others—is commonly assessed in collaborative contexts. In constructive collaborations, all teammates contribute equally toward the team's efforts and social loafing is minimized. Past work to explore this construct has sought to understand precisely how an individual's output changes as a function of being a member of a team in controlled laboratory situations (Latané et al., 1979). However, to estimate the extent to which social loafing impacts teams working in dynamic, naturalistic collaborations, metrics need to be extracted from task artifacts, as to not interfere with the team's performance. In the present study, we utilize task transcripts to understand the extent to which a single teammate contributes toward the team's CPS efforts. Specifically, we propose an objective measure of individual influence, beyond just an individual's proportion of the team's overall speech activity.

[1]School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA
[2]Human Systems Engineering, Arizona State University, Mesa, AZ, USA
[3]Institute of Cognitive Science, University of Colorado at Boulder, Boulder, CO, USA

Corresponding Author:
Julie L. Harrison, School of Psychology, Georgia Institute of Technology, 654 Cherry St NW, Atlanta, GA 30332, USA.
Email: julieharrison@gatech.edu

The present study builds on past work from the team sciences literature to provide initial validation for an objective metric that quantifies *CPS influence*—the extent to which each individual's speech activity contributes toward the team's socio-cognitive processes. Given that CPS is an interactive process, much of team coordination and problem solving occur via verbal communication channels. Thus, communication artifacts are a fruitful resource for modeling collaboration (Cooke et al., 2013; Hesse et al, 2015). While past work similar to the present study has relied on modeling only turn-taking behavior to understand an individual team member's influence (Gorman et al., 2020), we extend this work to the assessment of communication *content* during collaborations. Given the common practice of coding CPS communications for content (i.e., the presence of socio-cognitive processes), this was a natural extension of Gorman and colleagues' (2020) approach to the CPS setting.

Current Study & Hypotheses

We utilized existing data from a triadic task—originally reported by Amon et al. (2019)—wherein participants worked together to complete code.org's Minecraft *Hour of Code* (https://code.org/Minecraft) using videoconferencing. One participant was selected at random to serve as the controller (i.e., the only person who engaged with the virtual task interface directly) and shared their screen with the other teammates who contributed to the solution. This heterogenous but interdependent role manipulation—common to CPS tasks (Graesser et al., 2018)—allowed for a systematic comparison between the controller participant and the two non-controller participants. This design supported the validation efforts for the proposed metric, as the controller was uniquely positioned to have more influence over the team's output than the non-controller participants. Thus, if influence, as conceptualized in the current metric, were higher for the controller, we would have initial evidence in support of the novel measure.

Accordingly, we predicted that the controller's speech activity would have more influence on the team's CPS interaction than either of the other two team members. We also predicted that CPS influence may be related to learning outcomes and task performance, wherein those with higher influence (presumably, the controller) would have higher content knowledge and task performance. Further, we expected participants' subjective impressions of the collaborative session to be related to influence, so we included this as an exploratory analysis.

Method

Participants

As reported by Amon et al. (2019), 111 participants (63% female; M_{age} = 19.4) were recruited from a private Midwestern university and compensated with course credit. The 37 triads were formed based on schedule availability, where 19 participants from ten teams indicated familiarity with one or more teammates. Participants did not have prior computer programming experience.

Procedure

Participants were located in separate rooms, each equipped with a videoconferencing-enabled computer with a webcam and microphone, where they interacted via Zoom. One randomly assigned participant controlled interactions with the task interface and shared their screen. The other two participants engaged with the task by verbally communicating their ideas to help with planning and task execution. Audio, video, and screen activity were recorded.

Participants individually completed a demographic survey, as well as self-report scales (e.g., the Big Five Inventory) not reported here. Prior to beginning the experimental session, groups completed a 20-minute familiarization session of five easy levels and viewed three introductory computer programming videos that covered concepts such as loops and if statements. The experimental task had the groups complete a challenging programming task in *Hour of Code* (Figure 1), where they had 20 minutes to construct a 4 x 4 brick building using at least one if statement and loop. The task was further constrained by requiring that three bricks be constructed over water and only allowing the code to consist of 15 blocks or less.

Following the task, participants individually completed a survey indicating their subjective perceptions of the collaboration using Likert scale ratings of the following example statements: "I am satisfied with how we communicated with each other", and "I am satisfied with how we cooperated to complete the lessons." Lastly, participants individually completed a post-test to capture how well they understood the coding concepts utilized in the task. Because the data were collected over two semesters, where a 5-minute times-up warning was given only in the second semester, all outcome measures were Z-scored by semester.

Data Coding

Team-level task performance was scored from 0 to 5 by two independent raters on the five task requirements (e.g., use a loop, use 15 blocks of code or fewer, etc.). All discrepancies between raters were reconciled (Amon et al., 2019).

The communication content codes proposed by the Generalized Competency Model for CPS (Sun et al., 2020) were applied to the duration of the 20-minute interaction. Five of the 37 teams were excluded from behavior coding as one or more recorded channels were missing (audio, video, or screen recording). As shown in Table 1, content codes belonged to one of three CPS facets and noted the specific socio-cognitive behavior exhibited at the utterance

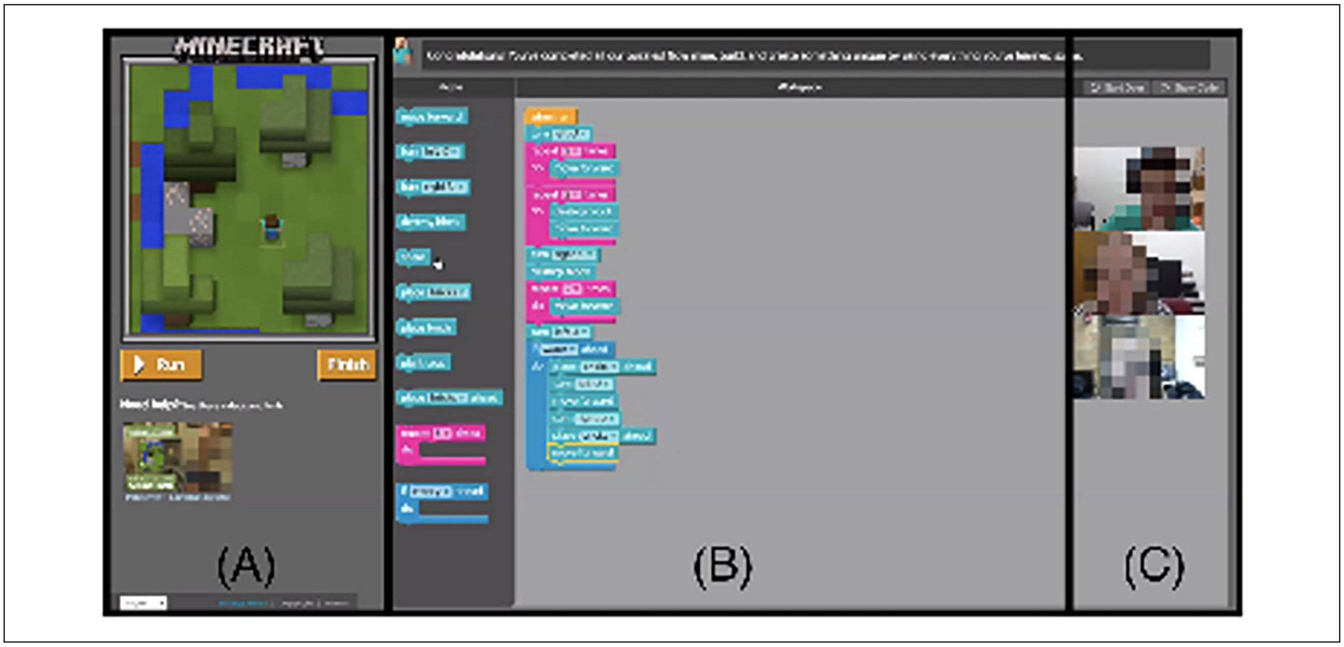

Figure 1. Screenshot of the *Hour of Code* environment (Amon et al., 2019). Blocks of code (B) were moved to produce solutions and solutions were tested in the Minecraft window (A). Teammates could also view each other's videos (C).

Table 1. CPS facets of the Generalized Competency Model adapted from Sun et al. (2020). Percentages represent the proportion of utterances ascribed to each facet.

Facet	Description	Example Utterance-Level Indicator
Constructing Shared Knowledge (33.0%)	The sharing of new ideas and the effort to understand them	Proposes specific solutions
Negotiation and Coordination (15.3%)	The process involved in reaching an agreed solution	Provides reasons to support/refute a potential solution
Maintaining Team Function (9.9%)	The process involved in sustaining the team's dynamic	Asks if others have suggestions

(i.e., single speaking turn) level. If the utterance did not contain evidence of a CPS behavior, no code was applied to that line in the transcript. If the utterance contained evidence for more than one CPS behavior, it could have been coded as belonging to more than one CPS facet. Two trained coders reached indicator-level reliability of 0.97 (Gwet's AC1) on two 5-minute video samples. They then each coded a randomly assigned half of the 32 videos (Amon et al., 2019). Verbosity (total words spoken captured from the IBM Watson Speech to Text service; IBM, n.d.) was also extracted from the transcripts as a content-free measure of team-member contributions.

Measures

To determine the extent to which any given team member's speech influenced the CPS communication content codes exhibited at the team-level, we used average mutual information (AMI; Abarbanel, 1996; Cover & Thomas, 2006). This metric has been used previously to determine which individuals' behaviors drove team-level adaptations (Gorman et al., 2020).

Two time series (X, Y) are input into the AMI function (Equation 1). If, in the symbolic time series, a symbol from X is independent from a symbol from Y, where $P_{XY}(x, y) = P_X(x)P_Y(y)$, then mutual information is 0. AMI, then, is the average of these probabilities over all symbols. AMI is theoretically greater than or equal to 0, as it is from the information theoretic approach and measured in bits, but its values are in reality constrained by the amount of information in the system. As the time series become more dependent, wherein by knowing one time series we know more information about the other, AMI becomes larger.

$$I_{XY} = \sum_{x_i y_i} P_{XY}\left(x_i, y_j\right) \log_2 \left[\frac{P_{XY}\left(x_i, y_j\right)}{P_X\left(x_i\right) P_Y\left(y_j\right)} \right] \tag{1}$$

Applied in the current study, we utilized the start and stop times of the utterances noted in task transcripts to create time series at both the individual- and team-levels. The individual-level time series represented whether (1) or not (0) that individual was speaking during a given 10 ms of the transcript. The team-level time series was constructed from the communication content codes exhibited by all participants on the team. For example, if a participant uttered a statement from 3.34 to 3.63 s, coded as *constructing shared knowledge*, the team-level time series would display 100 (the constructing shared knowledge nominal symbol) from the 334th cell to the 363rd cell, and that individual's time series would display a 1 in corresponding cells. AMI was calculated using the MATLAB AMI function (Shrestha, 2005) with bin size 10 and a lag of 0. The metric was calculated once for each participant in a team, wherein a participant's time series (X) was assessed for the AMI it shared with the team's time series (Y). This allowed us to understand which participant's communications accounted for the most mutual information, and thus influence, of the team's CPS-relevant utterances.

We also calculated the ratio of an individual's CPS verbal behaviors to their team's total number of utterances to understand how this simple proportion would covary with the novel influence metric. We refer to this ratio as *proportion of verbal behaviors*. For example, if Participant A contributed 30 utterances overall, but only 4 of those were coded as containing evidence of CPS processes, and their team's transcript was 100 utterances long, Participant A's proportion of verbal behaviors would equal 0.04. This proportion was calculated using inputs analogous to the AMI measure of influence so we included this in addition to verbosity as it may share more variance with influence than the simple verbosity measure.

Results

All analyses were performed in R (R Core Team, 2022) with comparisons between the controller and non-controller participants, wherein individual-level measures and covariates—influence, post-test scores, subjective ratings, verbosity, and proportion of verbal behaviors—were averaged for the two non-controller team members. Comparisons between the controller and the averaged non-controller participants were done via linear mixed effects models using a dummy coded variable with the controller entered as the reference group. Team membership was included as a random intercept.

First, we regressed influence on role with verbosity and proportion of verbal behaviors as covariates (Equation 2). While controlling for these factors, results suggest that the controller participants had significantly more influence over the CPS interaction than the non-controllers, which provides support for our hypothesis and initial validation for the AMI metric, $\beta = 0.15$, 95%CI [0.02, 0.06], $p < .001$. Said differently, the controller's speech had significantly more dependency with the entire team's CPS codes than the other two participants. With respect to covariates, those with higher verbosity tended to have higher influence scores, $\beta = 0.14$, 95%CI [0.08, 0.20], $p < .001$, but proportion of verbal behaviors was not significantly related to influence, $\beta = -0.04$, 95%CI [-0.10, 0.02], $p = .20$).

$$\text{lmer}\left(\begin{array}{l}\text{Influence} \sim \text{Controller}\left[\text{Y,N}\right] + \\ \text{Verbosity+Proportion Verbal Behavior} \\ + \left(1|\text{Team Number}\right)\end{array}\right) \quad (2)$$

We then sought to understand how influence related to post-test scores. We retained participant role as a predictor and the covariates from Equation 2, however in this second model influence was entered as a predictor and post-test score as the outcome variable. Neither predictor, influence, $\beta = 0.05$, 95%CI [-0.24, 0.35], $p = .72$, nor, participant role, $\beta = -0.14$, 95%CI [-0.63, 0.36], $p = .58$, were significantly related to post-test performance. Thus, participants appeared to retain task concepts regardless of their role or influence on CPS codes. Further, neither of the covariates related to task performance (verbosity, $\beta = 0.15$, 95%CI [-0.16, 0.46], $p = .33$; proportion of verbal behaviors, $\beta = 0.01$, 95%CI [-0.31, 0.33], $p = .93$). While a potential explanation for these null findings is provided later, the results of this model did not support our hypothesis that influence would be related to performance.

Next, we ran a model to understand how influence and participant role predicted subjective ratings of the collaboration. Similar to findings from post-test scores, neither the predictors (influence, $\beta = -0.05$, 95%CI [-0.31, 0.20], $p = .68$; participant role, $\beta = 0.19$, 95%CI [-0.37, 0.75], $p = .51$) nor covariates (verbosity, $\beta = -0.10$, 95%CI [-0.42, 0.21], $p = .51$; proportion of verbal behaviors, $\beta = -0.17$, 95%CI [-0.50, 0.16], $p = .32$) significantly related to participants' subjective ratings of the collaboration.

We then ran correlations to understand how the participant-level variables in the above models were related. Only verbosity and proportion of verbal behaviors were significantly correlated, $r(62) = 0.59$, 95%CI[0.40, 0.73], $p < .001$. Pairwise correlations between the verbosity and proportion of verbal behaviors covariates and the other variables—influence, subjective ratings, post-test scores—were all non-significant, $p > .05$.

Lastly, we assessed the effect of role-specific influence on the team-level measure of performance. Because controller and non-controller influence between teammates was highly correlated, $r(30) = 0.97$, 95%CI[0.93, 0.98], $p < .001$, we ran two separate linear models to assess the effect of influence on task performance. While both models were not significant, controller influence, $\beta = 0.11$, 95%CI[-0.26, 0.48], $p = .55$, was a relatively better predictor of team-level task performance than non-controller influence, $\beta = 0.06$, 95%CI[-0.31, 0.44], $p = .73$.

Discussion

The present study provides initial validation for a novel measure of an individual's influence on team-level CPS processes. The metric utilized communication content codes in the adaptation of previously explored team science methods to bridge an individual's contributions to output at the team level (Gorman et al., 2020). Given the results of the first model, where the controller had more influence over the team's CPS communications than the other participants, we can conclude that the influence metric, as operationalized using AMI, does capture an individual's contributions to collaborative interactions. The manipulation to participant roles allows us to make this casual conclusion, as the controller was uniquely positioned to have more influence than the non-controller participants. Though influence was significantly related to verbosity in this model, it may not always be the case that an individual's high word count is influential to CPS processes. In future efforts, we aim to investigate influence in other contexts and iterations to understand whether the metric can capture instances of substantial contributions independent of verbosity.

While individual-level post-test scores were hypothesized to relate to influence, it is not entirely unexpected that the two measures were unrelated, as they each may account for different aspects of collaboration. It is likely that the influence metric captures aspects of the CPS interaction that fall along the *social* dimension of this socio-cognitive activity. Social aspects of CPS include the behaviors it takes to coordinate and regulate the demands of a team's collaborative efforts. This may involve such activities as ensuring all members have a turn or chance to contribute (Hesse et al., 2015), a quality that would in part be captured by the present study's conceptualization of influence. On the other hand, the individual post-test scores in the present study captured the *cognitive* gains of engaging with the activity—in this case, computer programming concepts. Thus, whether in the controller role or not, participants gained computer programming knowledge. We also investigated team-level performance as it relates to role-specific influence scores. Our hypothesis that influence is related to task performance was partially supported given the controller's influence was more related to this team-level performance score than the non-controllers' influence.

The final mixed-effects model analyzed to what extent subjective ratings of the collaboration would relate to influence and participant roles. Based on the items in this scale, it is not surprising that these constructs were not related. The scale asked participants to respond to prompts that inquired about the interaction in general, not their perceptions of how they themselves contributed to the group. Future efforts will address this limitation via self-report measures that more centrally inquire about how participants felt their individual contributions drove the group's problem solving processes and dynamic.

More specifically, future studies will assess the degree to which the influence metric correlates with social loafing and emergent leadership behaviors. In classroom learning, social loafing has been measured using self-report scales (e.g., Linnenbrink-Garcia et al., 2011) and behavior coding (e.g., Nieswandt et al., 2020), however additional measures that objectively assess this construct would be useful for applications like teacher dashboards. It is currently envisioned that this metric could identify which group member(s) have the least influence in CPS discussions, should the task not involve clearly defined roles, which may be a marker of social loafing. However, social loafing may also be imposed on an individual through forces extrinsic to the classroom interaction (e.g., existing power dynamics between students) so teachers and researchers should use discretion when interpreting the measure. Further, emergent leadership, which occurs when leadership status is not based on designated roles but instead spontaneously arises, may also be explored within this measure, where high influence scores may correlate with leadership behaviors (Hollander, 1960; Yoo & Alavi, 2004).

In all, the influence metric allows us to assess the degree to which CPS interactions are distributed equally amongst the group by assessing the range of influence scores within a team. If the team is indeed collaborating absent of specified roles, and the range of influence is close to 0, the team could be understood to have *equitable interaction*. However, this metric should and will be applied cautiously, as not all students contribute to collaborations verbally—the modality considered in the present study. To truly understand a team member's contribution, more accurate measures will include multiple modalities. In future work, we will explore cross-modal effects, such as whether an individual's gaze behavior carries mutual information with team communication streams.

Future Directions: Classroom Data & Real-Time Modeling

We ultimately intend to assess influence in real-world classroom collaborations. Classroom data is inherently noisier with generally less intelligible audio and often inconsistent group sizes and roles, thus our validation effort first focused on controlled lab data. As an initial effort, we have also computed the measure on excerpts of a dataset of transcribed videos from dyads and triads of middle school students collaborating on a task that required them to learn about sensors through hardware and block programming. These transcripts currently have only 5-minute snippets of each team's transcript coded for CPS content using the Generalized Competency Model from Sun et al. (2020).

We calculated the influence measure for five of these 31 5-minute transcripts, selected at random. We report on the results of one team to demonstrate how the influence metric reflects classroom interactions. This dyad displayed an unequal distribution of influence, where Student 1's influence score was 0.04 and Student 2's was substantially larger at 0.24. As verbosity was related to influence in the Minecraft task, the two measures remain related in this classroom example, where Student 1 uttered 34 words and Student 2 uttered 86 words. The difference between the two metrics is that influence is computed as a kind of nonlinear correlation between speech activity and CPS behaviors, whereas verbosity provides a more direct measure of speech. Further, as can be seen in the excerpt from their transcript below, Student 2's utterances were more substantive and drove their solution forward with directives more so than Student 1's speech. This pattern persisted throughout the 5-minute interaction.

> Student 2: Cause I don't think ours is working.
> Student 2: I guess we could unplug it and plug it back in
> Student 1: We could?
> Student 2: And then do you want to plug it back in?
> Student 2: And then we just plug it.
> Student 2: Is this put on right?
> Student 1: I thought so.
> Student 2: Maybe wait for this to download and then. . .
> Student 2: Oh, it works now.

While one virtue of the influence measure is its consideration of both individual- and team-level data, this measure was implemented specifically for future use as a real-time (i.e., dynamic) measure of CPS influence. Moving window approaches to AMI have been applied in other team settings to detect critical events in the environment using diarized speech data (i.e., who is speaking and when), devoid of communication content information. Gorman et al. (2020) report on their success modeling individual and team communications in real time to recognize events using turn-taking behavior alone. This method enabled them to also determine which members drove the team's response to the event. In future efforts, we will adapt the Gorman et al. (2020) approach and apply it to classroom collaborations, making use of communication content codes, where the critical events we aim to recognize are such things as moments of insight or points at which the teacher provides guidance to the group.

Conclusion

In the present study we partially validated a multilevel measure intended to quantify an individual's verbal influence toward their team's CPS processes. While this metric has utility as-is, future work will systematically explore its validity in classroom collaborations and as a dynamic measure for potential use in collaboration dashboards and assessments.

Acknowledgments

This research was supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not represent views of the NSF. The authors also acknowledge the data collection efforts of iSAT members and the research assistants who aided in data coding.

References

Abarbanel, H. (1996). *Analysis of observed chaotic data*. New York, NY: Springer.

Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science*, *43*(10), e12787.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, *37*(2), 255-285.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: John Wiley.

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human factors*, *62*(5), 825-860.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest*, *19*(2), 59-92.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In *Assessment and teaching of 21st century skills* (pp. 37-56). Springer, Dordrecht.

Hollander, E. P. (1960). Competence and conformity in the acceptance of influence. *The Journal of Abnormal and Social Psychology*, *61*(3), 365–369. https://doi.org/10.1037/h0049199

IBM. (n.d.). IBM Watson Speech to Text service. Retrieved from https://www.ibm.com/watson/services/speech-to-text/. Accessed: 2018-05-02.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*(6), 822–832. https://doi.org/10.1037/0022-3514.37.6.822

Linnenbrink-Garcia, L., Rogat, T. K., & Koskey, K. L. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psychology*, *36*(1), 13-24.

Nieswandt, M., McEneaney, E. H., & Affolter, R. (2020). A framework for exploring small group learning in high school science classrooms: The triple problem solving space. *Instructional Science*, *48*(3), 243-290.

Organisation for Economic Co-operation and Development. (2017a). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving* (Rev. ed.), Retrieved from the OECD website: http://dx.doi.org/10.1787/9789264281820-en

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shrestha, D.L. (2005). AMI. MATLAB, MathWorks. https://www.mathworks.com/matlabcentral/fileexchange/7936-ami-and-correlation

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

Yoo, Y., & Alavi, M. (2004). Emergent leadership in virtual teams: what do emergent leaders do?. *Information and organization*, *14*(1), 27-58.