Exploring Linguistic Indicators of Social Collaborative Group Engagement (Manuskript)

Loris T. Jeitziner, Lisa Paneth, Carmen Zahn
loris.jeitziner@fhnw.ch, lisa.paneth@fhnw.ch, carmen.zahn@fhnw.ch
University of Applied Sciences and Arts Northwestern Switzerland

Dirk U. Wulff
wulff@mpib-berlin.mpg.de
Max-Planck-Institut für Bildungsforschung

Abstract: The recent advantages in methodological development renders it possible to automatize rating procedures, that have been carried out by human raters. Ongoing development of a combinations of NLP methods to assess Social Collaborative Group Engagement – in CSCL-learning groups – is presented. Preliminary results suggest that Social Engagement may be assessed through Sentiment, Group Cohesion and the Uniformity of participation in verbal communication through linguistic indicators in verbal communication.

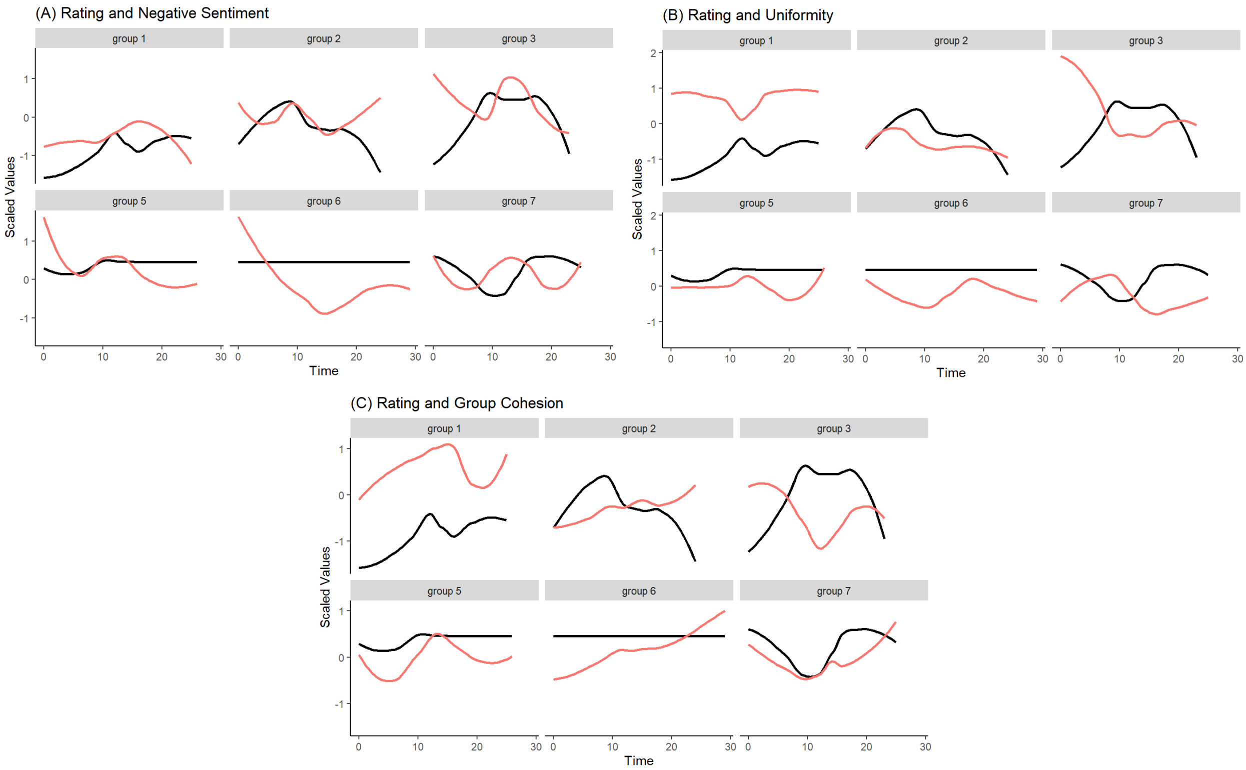
Introduction

Research in *Computer-supported collaborative learning* (CSCL) is inherently striving for further methodological advances. Technological innovations, such as faster computational processing and accompanying methodological usage, enable us to analyze large quantities of data. Thus, new approaches can and should be explored.
A promising approach is the computational analysis of text, or in the case of CSCL, verbal communication in the form of transcripts. Concerning the analysis of group processes in CSCL, computational analysis of verbal communication is already widely applied (Cress et al., 2021) and suggested for further application (Wise et al., 2021). However, regarding the construct of collaborative group engagement, no advances have been made, yet. Therefore, the questions arises, whether collaborative group engagement in CSCL-learning groups may be assessed through linguistic indicators in verbal group communication using *Natural Language Processing* (NLP). In this study, the progress of the development of a method to assess the dimension of social engagement is presented.

Methods

This preliminary method was developed with data of an exemplary observational study of CSCL-learning groups. In this study, N = 20 students were working in N = 6 groups (of 3 to 4 members) in virtual groups in an online learning setting. The groups were tasked with solving a hidden profile assignment via a video conference tool. Participants were provided with shared and unshared information about a fictional "murder case" and then given 45 minutes to reach consensus and find the "murderer". To assess a baseline measure of Social Engagement, an existing rating scheme was applied (Sinha et al., 2015).
To assess social engagement an observation based method was adapted - originally developed by Sinha et al. (2015) for capturing three levels of social engagement (high, moderate, low). The method consists of segmenting observations of videotaped student groups into one-minute intervals and rating these segments. The rating is based on judgments regarding the learning groups participation uniformity, group cohesion and sentiment of communication. A computational method was developed to measure the groups (1) participation uniformity, (2) group cohesion (responsiveness) and (3) negative sentiment based on linguistic indicators. Transcripts of each learning groups verbal communication were used to develop the following methods. (1) To capture the groups participation uniformity, each participants words were counted over each 1-minute sequence of the group task. The words counts then were put in relative proportion to the sum of the groups word count in this sequence. The resulting relative proportions then were used to calculate variances in order to get the groups participation uniformity. (2) To capture group cohesion (responsiveness), *Latent Semantic Analyis* (LSA) was applied to extract the semantic similarity of each participants responses in a 1-minute sequence. (3) To capture the negative sentiment of the learning groups communication, lexica-based sentiment analysis was applied. Each word in the verbal communication (that existed in the sentiment lexica) was assigned with a sentiment score that indicated valence and strength of the sentiment. The scores were then aggregated that resulted in sentiment scores over each 1-minute sequence. Subsequently, the three methods were descriptively compared with the baseline ratings of social engagement. All calculated values were smoothed over time for the descriptive comparison.

Figure 1
Smoothed and scaled Social Engagement Ratings (black) over the course of the task, compared with smoothed and scaled Negative Sentiment, Group Cohesion and Uniformity (each in red).



Exemplary Results

The results indicate, that for each of the three methods (i.e. Negative Sentiment, Group Cohesion & Uniformity) compared with the baseline measure (i.e. rating of Social Engagement), there are noticeable associations.
In Figure 1 – A, the negative sentiment seems to align with the Social Engagement ratings. In other words, the more the groups verbal communication was negative, the lower the groups were rated on Social Engagement.
In Figure 1 – B, the Uniformity seems to negatively relate to the Social Engagement ratings. The more the conversation distribution of the groups were uniform (indicated by lower value), the higher each groups were rated on Social Engagement. In Figure 1 – C, the Group Cohesion seems to positively relate to the Social Engagement ratings. The more the groups indicated a high group cohesion, the higher each groups were rated on Social Engagement (except group 3).

Discussion

The goal of this presented study is the assessment of collaborative group engagement with linguistic indicators from transcripts of CSCL-learning groups. Three common NLP methods were used to compare them with ratings of Social Engagement. The visualization indicate noticeable associations between the manual ratings of Social Engagement and measured Sentiment, Group Cohesion and Uniformity of participation in the verbal communication. Further development of the method deems promising. In future research, the ratings and linguistic indicators should be modelled with inferential statistics to test the hypothesized association. A machine learning approach to train and test an algorithm that automatically rates verbal communication on Social Engagement is feasible. Further dimensions of collaborative group engagement (e.g. behavioral & cognitive) may be included to achieve a more comprehensive picture.

References

Cress, U., Rosé, C., Wise, A. F., & Oshima, J. (Eds.). (2021). *International Handbook of Computer-Supported Collaborative Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 273–307. https://doi.org/10.1007/s11412-015-9218-y

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In *International handbook of computer-supported collaborative learning* (pp. 425–443). Springer.