Exploring Socially Shared Regulation with an AI Deep Learning Approach Using Multimodal Data

Andy Nguyen, University of Oulu, Andy.nguyen@oulu.fi
Sanna Järvelä, University of Oulu, sanna.jarvela@oulu.fi
Yansen Wang, Microsoft Research Lab (MSRA) Shanghai, yansenwang@microsoft.com
Carolyn Rosé, Carnegie Mellon University, cp3a@andrew.cmu.edu

Abstract: Socially shared regulation of learning (SSRL) is essential for the success of collaborative learning, yet learners often neglect needed regulation while facing challenges. In order to provide targeted support when needed, it is critical to identify the precise events that trigger regulation. Multimodal collaborative learning data may offer opportunities for this. This study aims to lay such a foundation by exploring the potential for using machine-learned models trained on multimodal data, including electrodermal activities (EDA), speech, and video, to detect the presence of SSRL-relevant process-level indicators in successful and less successful groups. The study involves thirty groups of secondary students (N=94) working collaboratively in five physics lessons. Considering the demonstrated positive results of machine-learned models, the advantages and limitations of the technical approach are discussed, and further development directions are suggested.

Introduction

Self-regulated learning theory describes the phases and processes through which learners navigate towards achieving learning goals in the face of challenges posed across situations. Collaborative learning has advanced self-regulated learning theory contextualized within collaborative contexts under the distinct headings of co-regulation and socially shared regulation (Hadwin et al., 2018). While self-regulated learning is important for individual academic achievements, previous studies have shown that processes identified as socially shared regulation of learning (SSRL) enhance group performance and produce better learning outcomes during collaborative learning (Isohätälä et al., 2017; Järvelä, Järvenoja, et al., 2019). Nevertheless, self-regulated learning is often difficult and not always conducted by the learners while confronting learning challenges. Regulation of learning is even more difficult in collaborative learning when the individual processes must be coordinated and communicated between learners within a group. Accordingly, it is suggested that collaborative learning interventions should aim to promote SSRL.

However, it is difficult to provide just-in-time support for SSRL since the cognitive and affective processes at its core are unobservable by learners and, thus, they are not always aware enough in order to regulate them (Azevedo & Gašević, 2019; Järvenoja et al., 2018). Research in self-regulated learning has highlighted the methodological challenges in capturing the dynamic and cyclical nature of regulation. Although extensive research has been published examining self-regulated learning and its influence on learning success, a systematic understanding of how SSRL evolves in collaborative learning is nascent. While self-regulation of learning is difficult to capture at the individual level, it is even more difficult to detect and support SSRL in the face of the dynamic interactions and influences among learners in collaborative learning settings. If there would be an efficient and effective approach to capturing regulation in collaborative learning, it could significantly progress research on SSRL.

Recently, initial methodological progress towards establishing a means to make SSRL observable has been published (Hadwin, 2021; Järvelä, Järvenoja, et al., 2019). In particular, Dindar et al. (2020) has utilized electrodermal activity (EDA) data to investigate the interplay of temporal changes in self-regulated learning processes connected with behavior, cognition, motivation, and emotion in computer-supported collaborative learning. The relative importance of multimodal data analysis on SSRL research has also been subject to considerable discussion. However, despite the potential to advance domain knowledge and offer innovative learning support, recent studies have recognized issues and challenges that still need to be addressed using multimodal and multichannel data regarding self-regulated learning (Azevedo & Gašević, 2019). One of the most frequently stated problems relates to integrating the plethora of theoretical models and frameworks from multiple disciplines including educational and computational sciences. Consensus built from consideration of recent studies calls for design and application of theoretically based and empirically derived approaches for collecting, measuring, and modeling multimodal and multichannel data to elucidate the complexity and temporality of underlying processes to address learners' self-regulatory needs (Azevedo & Gašević, 2019; Hadwin, 2021). Considering recent methodological progress and related data for SSRL, advanced technologies under the headings of Artificial Intelligence (AI) and machine learning could be further harnessed in order to push forward the nascent science to this end.

In this paper, we present an interdisciplinary study attempting to address these methodological challenges and issues in utilizing advanced technologies on multimodal and multichannel data to study self-regulated learning. This study aims to explore different types of interactions for learning regulation between successful and less successful groups in collaborative learning. We focus more specifically on designing a machine learning model with multimodal data to automatically detect these regulatory interactions. Our study addresses the following research questions:

1) How do interactions for learning regulation differ between successful and less successful groups in collaborative learning?
2) How can machine learning architectures and multimodal data representations be designed for the successful detection of regulatory interactions for both successful and less successful groups in collaborative learning?

This study offers some important insights into the application of advanced technologies in examining regulation in collaborative learning. The experimental work presented here provides one of the first investigations into how multimodal AI deep learning could be used to exploit SSRL with the fusion of features gathered from multiple data modalities and channels. Although the study does not yet examine the direct effects of SSRL on learning performance or specifically offer a prediction of the level of success for collaborative groups, it does present a compelling novel approach and meaningful evaluation of challenges and limitations to inform future work in this area within this research community. Furthermore, this study will contribute towards methodological progress to build a foundation for future theoretical investigation of SSRL and for the development of SSRL interventions.

Theoretical Approach – Regulation of Learning in Collaborative Learning

Our theoretical approach to the study is grounded in self-regulated learning (SRL) and socially shared regulated learning theory suggesting that regulation of learning is not solely a learning process at the individual level but also involves social and contextual processes (Hadwin et al., 2018).

Self-regulation of cognition, emotion, and behaviour has been a critical aspect of learning and successful academic performance (Winne, 2019). Self-regulation refers to the process in which a learner goes through goal-setting, self-monitoring, self-instruction, and self-reflecting to understand and manage their own learning. Research on self-regulation has progressed over decades with several theoretical perspectives and models proposed to describe and provision regulation in learning (Panadero, 2017). As the field has evolved, recent attempts have been made to investigate the phenomenon of regulation in a complex and dynamic collaborative learning environment. Going beyond self-regulation of an individual learner, co-regulation and socially shared regulation presented other social forms of regulation operating in collaborative learning. While co-regulated learning (CoRL) generally explains the learners' regulation with support from team members through social interactions, socially shared regulation of learning (SSRL) addresses the complementary perspective relating to a group's deliberate, strategic, and transactive planning, task performance, and reflection in learning (Järvelä, Järvenoja, et al., 2019). Although considerable effort has been made to understand the complex process of regulation of cognition, motivation, and emotion (Winne, 2019), the mechanisms that underpin regulation in collaborative learning are not yet fully understood.

Regulation of learning in collaborative settings is a complex and cyclical process in which the learners regulate themselves and each other (Järvelä, Järvenoja, et al., 2019). In the best cases, learners conduct all three forms of regulation together while confronting learning challenges progressing towards desired learning outcomes. These three forms of regulation in collaborative learning often occur alongside interactions aimed towards regulation (Isohätälä et al., 2017). Group interaction processes play a critical role in regulation promoting group success in collaborative learning. This study focuses on different types of interactions for regulation in collaborative learning. Previous studies have recognized a variety of specific challenges in capturing regulation in collaborative learning due to the complexity, "unobservability, temporality, multifaceted and situated nature of learning regulation" (Azevedo & Gašević, 2019; Hadwin, 2021; Järvelä, Järvenoja, et al., 2019). The methodological progress we report herein capturing regulation could offer a solid foundation for extending understanding regulation of learning, developing relevant theories, and designing support for promoting regulation of learning hence improving learning.

Methodological Approach – Multimodal Data and Machine Learning for Understanding Regulation of Learning

Recent progress in learning regulation research has increased the demands for innovative and robust methods in examining social aspects of learning regulation (Järvelä, Järvenoja, et al., 2019; Winne, 2019). Recently, research on learning analytics and educational data mining has exploited cutting-edge data collection and analytic technologies to reveal new insights into learning and learning processes with the goal of advancing learning sciences theories (Nguyen et al., 2017, 2020; Noroozi et al., 2019). Developments in sensor technologies have provided accessibility to different data channels such as physiological data. Collecting and analyzing data from multiple modalities and channels has provided exposure to signals thought to be related to regulatory processes treated previously as unobservable when the primary method for studying these processes was through self-report instruments such as questionnaires or interviews (Azevedo & Gašević, 2019; Järvelä & Bannert, 2019; Noroozi et al., 2019).

Previous studies have utilized multimodal data to provide and triangulate evidence of temporal and adaptive processes of learning regulation (Järvelä, Järvenoja, et al., 2019). Although the recent pioneers in using multimodal data for examining learning regulation have contributed to better understanding the complex nature of (S)SRL (See Järvelä & Bannert, 2019; Noroozi et al., 2019), they have encountered several issues related to synchronizing data from different sources (Azevedo & Gašević, 2019; Hadwin, 2021). Nevertheless, prior studies analyzing and triangulating multimodal data have provided valuable insights and progressed regulation concepts that provide the foundation for this work. The current study builds on and extends this foundation through the development of technological advances that enable effective integration of the synergistic features from multimodal data for better understanding and supporting learning regulation.

Advanced Artificial Intelligence (AI) technologies, specifically what is currently known as deep learning, have offered the capabilities to not only automatically detect cognitive and emotional activities in real-time but also objectively measure and classify relevant learning behaviors. For instance, the application of deep learning across different media has been applied to the problem of detecting human emotions including micro-expressions (Tzirakis et al., 2017). Furthermore, AI deep learning methods potentially allow the fusion of features obtained from multiple data modalities to enhance the accuracy of recognizing and classifying human cognitive and emotional activities (Tzirakis et al., 2017). As cognitive and emotional activities are at the core of learning and learning regulation, the use of multimodal AI deep learning methods offers the potential for gaining novel understandings of the phenomenon and providing effective support to promote SSRL. Despite the potential of advanced technologies such as AI deep learning in examining SSRL, there are only a few studies that have empirically investigated the phenomenon. Indeed, apart from the study proposed by Nguyen et al. (2021), our search of the literature failed to find any prior studies utilizing multimodal AI deep learning to examine interactions for regulation in collaborative learning. We thus argue that the contribution of the current study is in affording the potential to establish methodological foundations to push (S)SRL research forward.

Research Methods

Participants and Procedures

The participants of this study included secondary school students (N = 94) working collaboratively in groups of three or four during five physics lessons. They were divided into 30 groups and the group assignment was based on the grades they earned in previous science courses. Students and their processes of learning and collaboration were monitored and documented with video recordings and through individual-level physiological measures.

Data collected from different modalities and channels may allow for uncovering the complex interactions of both the visible actions and the invisible mental and metacognitive processes in collaborative learning (Azevedo & Gašević, 2019). Following this line of research, our study attempted to collect the multimodal data from several channels to reveal the interactions for regulation in groups. Particularly, we capture students' physiological activities through the measurement of Electrodermal Activity (EDA) by using Shimmer3 GSR + devices. These data reflect the physiological reactions that occurred during a learning situation, such as students' physiological activation related to emotion and cognition. In addition, we used Insta360 Pro video cameras and microphones to capture the visual and acoustic data reflecting engagement in learning activities in the classrooms' natural setting.

Qualitative Coding and Quantitative Data analysis

Qualitative coding through a multi-step analysis schema was conducted to label types of interactions for regulation. First, the videos and audio were divided into thirty-second segments. Previous research suggested that the time-based segmentation of videos for observing collaborative learning would benefit the analysis by enabling time series modeling of a progression over time and providing a consistent unit of analysis (Sinha et al., 2015). The types of interactions for regulation were classified using four separate labels: 1) metacognitive interaction, 2) socio-emotional interaction, 3) task execution and 4) other interactions. The identification of these types of social interactions would lead to the recognition of SSRL when it emerged as SSRL is socially situated and achieved through social interactions (Isohätälä et al., 2017). By examining the four types of social interactions for regulations, we could better understand how SSRL could be effectively developed to enhance group performance through different social interactions. The coding scheme is shown in Table 1. To ensure the reliability of the labels, we validate the coding by conducting interrater reliability testing on the initial set of coded data. This initial set of data are coded for ten videos randomly selected from the dataset. The result shows a substantial agreement among the results with Cohen’s Kappa values above 0.7.

Table 1: Qualitative Coding Schemes for Preparing Interaction Data

Type of Interactions for Regulation	Description of behavior	Examples
Socio-emotional interaction Socio-emotional interaction was coded when group members expressed clear indicators of positive/negative affect or made a positively/negatively charged comment. Emotional expressions included verbal or other clear indicators of positive or negative affect and negatively or positively charged interaction.	Verbal indicators, e.g.: - positive/negative content - positive/negative tone of voice - sarcasm - laughing, singing, groaning, whining Bodily indicators, e.g.: - smiling - dancing - sighing - facepalm Emotionally charged interaction, e.g.: - Joking, praising - Arguing, criticizing	*"This [shimmer] gets stuck everywhere. I hate these. I wish this would be over already."* *"You sound like a hamster!" *laughter** *"How are you able to draw such straight lines without a ruler? Look! She didn't even use a ruler to draw this line!"* *"Can he be quiet? The mic doesn't like it when he does that all the time"*
Cognitive interaction Metacognitive interaction was coded when group members' discussion targeted their cognitive processes.	Discussion regarding: - Task goals - Task understanding (e.g. checking task demands) - Prior knowledge - Resources (e.g. procedures, strategies) needed to solve the problem - Progress with the task (e.g. checking/evaluating progress) - Selecting strategies (e.g. help-seeking) - Quality of task solution (e.g. evaluation of the correctness of the answer) and overall performance	*"Okay, what are our goals?"* *"Isn't this the first task? About volume?"* *"No, it is this way, we did this yesterday in math."* *"I don't think that we have time to do that. [The teacher] is already writing down the homework"* *"Jane, is this [answer] good?"*
Task execution Concrete work towards task completion. No interaction required.	- Reading or discussing task material or content - Executing task-related measurement or calculation - Otherwise physically executing task experiment - Writing down calculations or answer	*"[Exercise] 2. d) How long does it take sunlight to reach the Earth?"* *"What was the answer to exercise 3?"*
Other interaction Non-task-related interaction	Discussion regarding: - Other school-related issues - Free-time activities - Research setting	*"Maybe [the teacher] was talking about who will be in our group during home economics class"*

Before conducting machine learning modeling for automatic detection of regulatory interactions, we investigated whether the interactions for regulation differed among successful and less successful groups. We conducted a Pearson Chi-square analysis of homogeneity of proportions to examine the differences in regulatory interaction frequencies. We hypothesized that the proportion of the types of regulatory interactions is different between successful and less successful groups. This offers empirical evidence that detection of this four-way distinction will provide leverage towards our future work automatically detecting the level of success that groups are likely to achieve as they are working, and using that detection to prompt needed interventions.

AI Deep Learning Modelling and Evaluation

The multimodal deep learning modeling approach is adopted for developing the predictive model to detect interactions for regulation. Different AI deep learning frameworks will be used to extract features from each data modality and the extracted features will be then fused using a multi-layer perception. In particular, the pre-trained 18-Layer Residual Network (ResNet-18) (He et al., 2015) was applied to encode the video stream into a sequence of features, referred to as a vectorized representation. The same method is utilized to encode the audio data into a vectorized form. Besides, the electrodermal activity (EDA) processing approach follows prior work in which EDA is also used in a deep learning framework designed for emotion classification (Ganapathy et al., 2020). EDA data were decomposed into the tonic component or skin conductive level (SCL) and the phasic component or skin conductance response (SCR). EDA features were encoded and vectorized using a convolutional neural network (CNN). A dense layer using a multi-layer perceptron with a binary cross-entropy loss function was applied to perform deep learning to identify patterns within the transformed feature encoding related to the prediction labels. We train three AI deep learning models using multimodal data from: 1) successful groups; 2) less successful groups, and 3) all groups. Figure 1 demonstrated our prototype design for AI deep learning for (S)SRL.

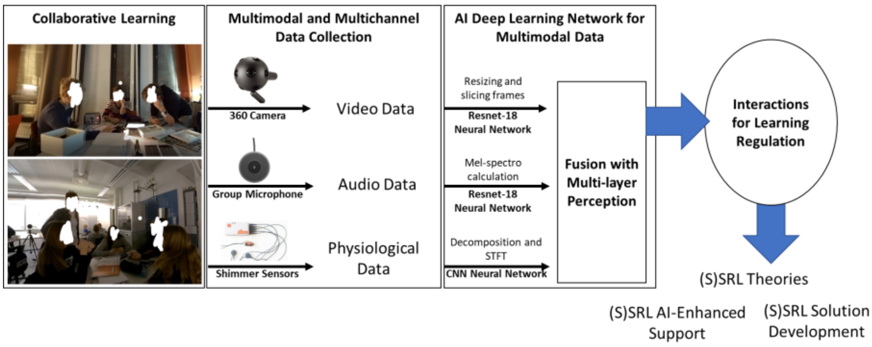

Figure 1: AI Multimodal Deep Learning for (S)SRL

For each AI deep learning model, we followed the typical machine learning experimental methodology to conduct training, validation, and evaluation. In particular, each multimodal dataset was randomly split into three subsets consisting of a training set, a validation, and a test set with a ratio of 8:1:1. The model is trained on the training set, where parameters are tuned to maximize performance on the validation set. The tuned model is then tested on the test set for evaluation. The common values of F1-score, mean Average Precision (mAP), and Accuracy score were obtained for the evaluation on the validation test and test set. The Area Under the Curve of Receiver Operating Characteristics (AUC-ROC or AUROC) is also computed for the evaluation of each model.

Results

We first address the question of the value of the four-way distinction used to code data segments to separate successful and unsuccessful groups, under the expectation that observations related to engagement in SSRL processes are included among the factors that separate these groups. The results of the Pearson Chi-Square test showed significance in the relationship between the observed distribution of types of interactions for regulation and group performance ($\chi^2 = 56.76$, $df = 3$, $p < 0.01$). In other words, the distribution of counts of different types of regulatory-relevant interactions is not independent of the success of the group. Next, we address the capability of the technical approach to automatically assign the four-way distinction to data segments. Table 2 demonstrated the results for AI deep learning evaluation for models on successful groups, less successful groups, and all groups.

Table 2: Evaluation Results for the AI Deep Learning Evaluation

	Successful Groups	**Less Successful Groups**	**All Groups**
Evaluation on the validation set			
F1-score	0.63	0.67	0.73
mAP	0.67	0.72	0.79
Accuracy	0.69	0.81	0.79
Evaluation on the test set			
F1-score	0.65	0.61	0.63
mAP	0.67	0.74	0.70
Accuracy	0.68	0.72	0.73

Overall, our results illustrate that using AI deep learning on multimodal data has the potential to automatically detect interactions for regulation in collaborative learning. In the evaluation on the validation set, we found that the predictive accuracy on detecting the regulatory interactions for successful groups (F1-score = 0.63, mAP = 0.67, and Accuracy = 0.69) is lower than for less successful groups (F1-score = 0.65, mAP = 0.67, and Accuracy = 0.68). The evaluation on the test set also confirmed that the deep learning model for successful groups (F1-score = 0.65, mAP = 0.67, and Accuracy = 0.68) is not as accurate as for less successful groups (F1-score = 0.61, mAP = 0.74, and Accuracy = 0.72). Nevertheless, the evaluation on both the validation set (F1-score = 0.73, mAP = 0.79, and Accuracy = 0.79) and evaluation set (F1-score = 0.63, mAP = 0.70, and Accuracy = 0.73) for the inclusive AI deep learning model using multimodal data from all groups has the capability to predict interactions for regulation at a reasonable level. Thus, we believe we have a proof of concept to drive forward research but note that in future research particular attention should be paid to performance specifically on unsuccessful groups.

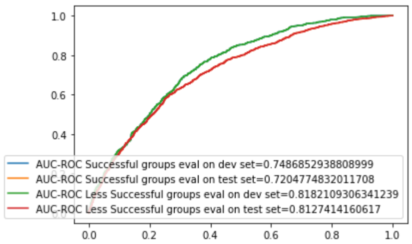

Figure 2: The Area Under the Curve of Receiver Operating Characteristics (AUC-ROC or

Figure 2 showed the results for the Area Under the Curve of Receiver Operating Characteristics (AUC-ROC or AUROC) for both models on the successful and less successful groups. The results show that both models could reasonably predict the interactions for regulation in collaborative learning (AUC-ROC > 0.72). The lower predictive power in the successful groups (AUC-ROC=0.74 and 0.72 for evaluation on the validation and dataset) could be explained by a higher complexity in interactions for regulations than the less successful groups (AUC-ROC=0.82 and 0.81 respectively). Following the approach by Fogarty et al. (2005), Chi-square test on the evaluation results confirmed the difference between two models with X^2 (1, N = 3408) = 29.36, $p < .001$. Future work could be conducted to examine this difference further. Nevertheless, the results have validated the feasibility of our model to recognize interactions for learning regulation in both conditions.

Discussion and Implications

The development of sensor-based technologies and machine learning techniques together presents a unique opportunity to leverage data collected from different modalities in the context of collaborative learning. Prior work has reported that multimodal data have shown promise for advancing our understanding of learning regulation (Noroozi et al., 2019), yet so far very little attention has been paid to how machine learning techniques could be implemented to examine the multiple facets and processes of learning regulation. This study explored the use of the AI deep learning approach on multimodal data to detect regulatory interactions for successful and less successful groups in collaborative learning.

First, we investigated whether the types of interactions for regulation are related to group performance. The results support our hypothesis that the proportion of types of regulatory interactions differs between successful and less successful groups. Previous research demonstrated that SSRL emerged from different types of interactions and contributed to the success of collaborative learning (Isohätälä et al., 2017). Therefore, the capabilities to capture regulatory interactions could establish a solid foundation for identifying triggers for regulation and providing support to promote SSRL, hence enhancing learning performance (Nguyen et al., 2021). This study extends the understanding of learning regulation by providing evidence for the difference in the proportion of types of regulatory interactions between the successful and less successful groups in collaborative learning. Accordingly, our results showed a significance for the difference in regulatory interactions and provisioned the importance of capturing different types interactions for regulation in collaborative learning.

The development and evaluation of AI deep learning models using multimodal data in this study demonstrated the feasibility of this approach to capture the regulation of learning in collaborative settings. Previous research has already recognized the potential of utilizing multimodal data for SSRL (Engelmann & Bannert, 2021; Järvelä, Malmberg, et al., 2019), but the field has struggled with an analysis of multimodal and multichannel data, even recognizing that the data allows for examining the interplay of different facets and processes of learning regulation (Hadwin, 2021). For instance, while Engelmann and Bannert (2021) analyses the screen recordings and verbal protocols to assess cognition and metacognition processes of regulation, Järvelä, Malmberg, et al. (2019) showed the use of facial expression extracted from video interactions and physiological data including heart rate and electrodermal activity (EDA) data to examine the SSRL facet of emotions. These studies provided evidence for the use of the multimodal approach for understanding the situated nature and temporality of regulation. Nevertheless, the synchronization of multiple data modalities across multiple data channels is complex and faces several challenges (Azevedo & Gašević, 2019; Järvelä & Bannert, 2019). Triangulating data from multiple channels could examine the interplay of multiple facets and processes of learning regulation and comprehend findings from each of the data sources (Järvelä, Malmberg, et al., 2019). While multimodal and multichannel data triangulation has provided valuable insights to extend our understanding of learning regulation, the cohesive multimodal analysis with potential synergistic effects for research on learning regulation has received scant attention in the literature. Since different data modalities from different channels reflected different facets and processes of learning regulation, utilizing synergistic features from multimodal and multichannel data could further address the temporality of regulation. However, the complexity and dynamics of synergizing features from multiple data modalities raised several methodological challenges. Therefore, this study demonstrated the opportunities offered by advanced technologies, specifically AI deep learning, for analyzing multimodal data in research on learning regulation.

This study has some limitations. First, the approach for qualitative coding and segmentation into thirty-second blocks is theory-grounded and supported by several published studies in the domain (e.g. Järvenoja et al., 2018; Sinha et al., 2015), yet other approaches should be examined and compared for improving performance. Second, while immersive video data hold valuable features for studying learners' behavior and learning, we noted that the current machine learning methods for video and image processing for this type of video data are still limited. Furthermore, the pre-trained models for analyzing video data are mostly trained with open databases in which the data are often retrieved from video scenes. The complexity and dynamics of authentic learning environments could influence the predicting performance of the pre-trained models. Thus, this study presented a unique contribution to the field with the model trained from a large dataset captured from real-life situations. And it could lead to a fruitful area for further work on SSRL theoretical investigation and multimodal methodological development.

Conclusion and Future Directions

In recent years, advanced technologies have brought opportunities for transformation across societal sectors, including education. Among these advanced technologies, Artificial Intelligence (AI) has been one of the most influential technologies and has been widely utilized in online education. However, there has been much less attention paid to the application of AI in face-to-face classroom settings. Although face-to-face learning has not been responsible for the bulk of available data from educational settings, the advancement of sensor technology and multimodal data collection methods has enabled the possibility of applying AI in the face-to-face classroom context. Despite the significant potential of AI in education, there remains a paucity of evidence on how AI could be applied to extend our understanding of learning in the classroom. This study demonstrated the implementation of an AI deep learning approach to address the methodological and theoretical challenges in the domain of self-regulated learning and collaborative learning. In particular, we designed and validated how the AI deep learning approach could reveal the previously "unobservable" processes of regulation by using multimodal data to automatically detect interactions for regulation. Future research could attempt to examine different AI approaches to offer novel methods to collecting and analyzing educational data, hence extending the methodological and theoretical boundaries in learning sciences. The results of this study also shed a light on the development of real-time support for regulation in collaborative learning. We look forward to future work building on and extending these findings in the design and creation of AI-enhanced solutions for promoting the regulation of learning.

References

Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2020). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*, *25*(3), 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*, *72*, 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Fogarty, J., Baker, R. S., & Hudson, S. E. (2005). Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. *Proceedings of Graphics Interface 2005*, 129–136.

Ganapathy, N., Veeranki, Y. R., & Swaminathan, R. (2020). Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features. *Expert Systems with Applications*, *159*, 113571. https://doi.org/10.1016/j.eswa.2020.113571

Hadwin, A. F. (2021). Commentary and future directions: What can multi-modal data reveal about temporal and adaptive processes in self-regulated learning? *Learning and Instruction*, *72*, 101287. https://doi.org/10.1016/j.learninstruc.2019.101287

Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 83–106). Routledge.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Computer Vision and Pattern Recognition*. http://arxiv.org/abs/1512.03385

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, *81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006

Järvelä, S., & Bannert, M. (2019). Temporal and adaptive processes of regulated learning—What can multimodal data tell? *Learning and Instruction*, 101268. https://doi.org/10.1016/j.learninstruc.2019.101268

Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological Progress in understanding socially shared regulation in learning. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 425–441. https://doi.org/10.1007/s11412-019-09313-2

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvenoja, H., Järvelä, S., Törmänen, T., Näykki, P., Malmberg, J., Kurki, K., Mykkänen, A., & Isohätälä, J. (2018). Capturing motivation and emotion regulation during a learning process. *Frontline Learning Research*, *6*(3), 85–104. https://doi.org/10.14786/flr.v6i3.369

Nguyen, A., Gardner, L., & Sheridan, D. (2017). A multi-layered taxonomy of learning analytics applications. *Pacific Asia Conference on Information Systems (PACIS) 2017 Proceedings*, 54–54. http://aisel.aisnet.org/pacis2017/54/

Nguyen, A., Järvelä, S., Wang, Y., Róse, C., Malmberg, J., & Järvenoja, H. (2021). Multimodal Deep Learning Model for Detecting Types of Interactions for Regulation in Collaborative Learning. *Proceedings of Annual Meeting of the International Society of the Learning Sciences (ISLS)*. https://2021.isls.org/proceedings/

Nguyen, A., Tuunanen, T., Gardner, L., & Sheridan, D. (2020). Design Principles for Learning Analytics Information Systems in Higher Education. *European Journal of Information Systems (EJIS)*. https://doi.org/10.1080/0960085X.2020.1816144

Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P. A., Juuso, I., & Seppänen, T. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*, *100*, 298–304. https://doi.org/10.1016/j.chb.2018.12.019

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. In *Frontiers in Psychology* (Vol. 8, Issue APR). Frontiers Media S.A. https://doi.org/10.3389/fpsyg.2017.00422

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

Tzirakis, P., Trigeorgis, G., Nicolaou, M. A., Schuller, B. W., & Zafeiriou, S. (2017). End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE Journal of Selected Topics in Signal Processing*, *11*(8), 1301–1309. https://doi.org/10.1109/JSTSP.2017.2764438

Winne, P. H. (2019). Paradigmatic Dimensions of Instrumentation and Analytic Methods in Research on Self-Regulated Learning. *Computers in Human Behavior*, *96*, 285–289. https://doi.org/10.1016/j.chb.2019.03.026

Acknowledgments

This work was funded in part by NSF grants IIS 1822831 and 1917955 and funding from Microsoft, Finnish Academy project no. 32438 and Oulu University Eudaimonia TRIGGER project funding.