

Article

Not peer-reviewed version

No pain no gain - Giving real-time emotional feedback in a virtual mirror improves collaboration in virtual teamwork

Nicklas Schneider , Ignacio Vazquez , [Peter A. Gloor](#) *

Posted Date: 4 June 2024

doi: 10.20944/preprints202406.0147.v1

Keywords: Virtual teamwork; real-time feedback; team performance; emotion recognition; videoconferencing; machine learning; virtual mirroring



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

No Pain no Gain - Giving Real-Time Emotional Feedback in a Virtual Mirror Improves Collaboration in Virtual Teamwork

Nicklas Schneider ^{1,2}, Ignacio Vazquez ³ and Peter A. Gloor ^{1,4*}

¹ Center for Collective Intelligence, Massachusetts Institute of Technology

² TUM School of Management, Technical University of Munich

³ System Design Management, Massachusetts Institute of Technology

⁴ Cologne Institute for Information Systems (CIIS), University of Cologne

* pgloor@mit.edu

Abstract: This study investigates the impact of real-time emotional feedback on the quality of teamwork conducted over videoconferencing. We developed a framework that provides real-time feedback through a virtual mirror based on facial and voice emotion recognition. In an experiment with 28 teams (84 participants), teams collaborated over Zoom to set up a virtual Mars station using custom simulation software. Participants were divided into 14 experimental teams which were shown the virtual mirror, and 14 control teams without it. Team performance was measured by the improvement in the Mars simulation output quality. Our analysis using correlation, multilevel regression, and machine learning revealed that fewer interruptions but an increasing number over time correlated with higher performance. Higher vocal arousal and happiness also enhanced performance. We confirmed that female presence in teams boosts performance. SHAP values indicated that high variability in happiness, head movement, and positive facial valence—an “emotional rollercoaster”—positively predicted team performance. The experimental group outperformed the control group, suggesting that virtual mirroring improves virtual teamwork and that interrupting each other more, while speaking less, leads to better results.

Keywords: Virtual teamwork; real-time feedback; team performance; emotion recognition; videoconferencing; machine learning

1. Introduction

In today's rapidly digitizing work environment, effective team collaboration is crucial for organizational success [1]. The shift towards virtual interactions, accelerated by the COVID-19 pandemic, has transformed occasional virtual meetings into daily necessities [2]. This transition underscores the need for research into the complexities of team dynamics and performance in virtual settings [3, 4]. Identifying key drivers of productivity in these environments is increasingly critical [5].

Historically, research on team collaboration and performance has relied on subjective measures such as surveys and questionnaires [6]. While valuable, these methods often fail to capture the real-time complexities of team interactions and are susceptible to biases [7]. This highlights the need for objective, real-time measures of team collaboration and performance [8], emphasizing data-driven approaches to understanding team dynamics.

Teamwork is critical in various settings, including schools, universities, and organizations, where collective efforts often yield superior outcomes compared to individual contributions [9, 10]. Effective teamwork hinges on clear goals, trust, role allocation, and conflict management [11]. Research has explored how teams form, develop, and achieve their objectives [12]. Understanding teamwork processes is essential for grasping how teams function and achieve goals. These processes convert inputs into outcomes through cognitive, verbal, and behavioral activities [13]. Team

processes are dynamic, adapting to task demands and environmental complexities, which influence team performance and success [14].

The effectiveness of teams focuses on efficiency, adaptability, and productivity within groups striving towards common goals [15]. Several models explain team effectiveness, including McGrath's Input-Process-Output (IPO) model, which systematizes teamwork into sequential stages: inputs, processes, and outputs [16]. The Input-Mediator-Output-Input (IMOI) model enhances the IPO model by recognizing the dynamic nature of teams through continuous feedback [17]. The Multiteam Systems (MTS) framework explores interactions among multiple teams within broader organizational structures, crucial for large projects [18].

2. Background and Motivation

The rise of information and communication technology has transformed traditional teamwork into virtual teamwork, especially post-COVID-19 [19]. Virtual teams collaborate primarily through electronic means, introducing challenges such as the absence of face-to-face interactions and managing cultural diversity [20, 21]. Effective management of non-verbal communication and technological disruptions is essential to maintain collaboration [22, 23].

Traditional teamwork models also apply to virtual teams. The IPO model, used in virtual team research, focuses on team dynamics and effectiveness in digital settings [24, 25]. The Virtual Team Maturity Model (VTMM) assesses virtual team progression through different maturity levels, identifying key competencies at each stage [26]. Media Richness Theory (MRT) and Social Presence Theory (SPT) emphasize selecting communication tools that effectively convey cues and enhance the sense of presence, essential for virtual teams [27, 28]. The Media Naturalness Theory (MNT) argues that deviations from face-to-face communication increase cognitive effort, impacting communication effectiveness in virtual environments [29].

Assessing teamwork effectiveness has evolved significantly. Historically, observational and subjective measures such as self-report questionnaires and direct observations were most prominent. Tools like the Teamwork Perception Questionnaire (TPQ) and Team Climate Inventory (TCI) captured individual perceptions of teamwork quality [30, 31]. Objective performance metrics linked teamwork to tangible outcomes like deadlines, error rates, and product quality [32].

Technological advancements have introduced new methods for measuring teamwork effectiveness. Simulation-based assessments present teams with predefined challenges to evaluate performance, communication, and adaptability [33]. Digital analytics leverage data from virtual platforms, analyzing communication patterns and interactions to gauge teamwork dynamics [34]. Recently, the analysis of body signals through computer vision and machine learning has emerged, offering real-time insights into team interactions, engagement, and emotional states [35, 36]. This leads us to explore the potential of predicting virtual team performance through multimodal data analysis, focusing on emotional states and conversational dynamics. Consequently, we propose our first hypothesis:

Hypothesis 1: Through an analysis of multimodal data, it is possible to accurately predict the teamwork performance of virtual teams.

Effective team performance is influenced by several key factors. Team composition, including a mix of skills, knowledge, and personalities, is critical, with diverse teams often balancing complementary skills and shared values [37]. Communication frequency and quality, along with robust feedback loops, are essential for effective collaboration [38]. Shared mental models among team members enhance coordination and reduce misunderstandings, leading to better performance [39]. Trust and a sense of belonging significantly contribute to a collaborative environment, enhancing overall team effectiveness [40]. Clear and aligned goals guide team direction and performance [41].

Emotions play a crucial role in team dynamics, influencing communication patterns, decision-making processes, and overall team cohesion [42]. Emotions, moods, and affects are fundamental to understanding team interactions. Emotions are intense, short-lived feelings triggered by specific

events, directly impacting team dynamics [43]. Moods are less intense but more enduring states that subtly influence team interactions and perceptions over time [44]. Affect encompasses both emotions and moods, providing a spectrum for analyzing emotional states in teams [45].

Positive emotional states, such as happiness, can improve performance, while balanced conversation dynamics characterized by equitable turn-taking enhance collaboration [46, 47]. This prompts us to explore the potential of predicting virtual team performance through multimodal data analysis, focusing on emotional states and conversational dynamics. This research builds on existing studies that highlight the significance of emotions and communication styles as significant drivers of team outcomes [48]. Given the insightful nature of emotions and conversational patterns in revealing team dynamics and effectiveness, we propose the following hypothesis:

Hypothesis 2a: Teams with higher levels of positive emotional states, such as happiness, will demonstrate improved performance [46, 47].

Hypothesis 2b: Balanced conversation dynamics of teams characterized by equitable conversation length and alternating turn-taking behavior among team members positively correlate with improved team performance [15].

Feedback is essential for team dynamics, serving as a critical tool for evaluating and improving performance [49]. Effective feedback in virtual teams can be categorized by its content, source, and level [50]. Content refers to the focus of the feedback, whether it is on team input, mediator processes, or output [51]. Feedback can be subjective, based on perceptions and evaluations, or objective, derived from concrete actions and events [52]. The level of feedback varies between individual, team, and team-plus-individual, providing insights into both collective and individual functioning [53].

Real-time feedback, with its immediacy, can significantly enhance learning and performance by providing instant reinforcement or correction [94]. For virtual teams, it is especially beneficial as it integrates performance-related information with team processes and psychological states, offering immediate insights and facilitating timely adjustments [50]. This immediacy addresses the lack of physical cues and direct interactions, fostering alignment and collaboration [95].

Hypothesis 3: Teams receiving real-time feedback about their emotional state will show improved performance, evidenced by increased performance scores [96].

One application of real-time feedback is virtual mirroring, where individuals are made aware of their emotions and communication behaviors in real-time [97]. This involves analyzing team communication patterns and mirroring them back to encourage introspection and adaptation, aligning behaviors with team goals. This fosters self-awareness and collective understanding of team emotions, optimizing team dynamics [98]. Use cases in large organizations have shown significant improvements in performance through this method [98], while respecting individual privacy by providing aggregated information. Considering the potential of real-time feedback in emotional regulation and its implications for team strategies, we ask: How does real-time feedback about team emotions affect the emotional state of the team? In response to the research question and based on existing findings in the literature, we have formulated the following hypotheses:

Hypothesis 4: Teams exposed to real-time emotional feedback will experience better emotional regulation, evidenced by reduced emotional variance [50].

Hypothesis 5: Teams receiving real-time feedback on their emotional states are likely to exhibit a broader emotional spectrum during collaboration, as indicated by higher variations of positive and negative emotions other than the neutral state [54].

This research aims to investigate virtual team dynamics and uncover the factors that contribute to team success in digitally transformed work environments. We address the overarching research question: *"How can real-time emotional feedback, informed by multimodal data analysis, enhance the performance and emotional dynamics of virtual teams, and what factors most effectively predict successful team collaboration?"* Utilizing advanced techniques such as multimodal data analysis and machine

learning, this study seeks actionable insights to optimize team performance with modern digital technology.

3. Materials and Methods

Building on existing research, our study investigates the impact of real-time emotional feedback on virtual team performance. We conducted a virtual field experiment with 84 participants from MIT (USA), Living Lab from the Institute for the Future of Education at Tecnológico de Monterrey (Mexico), and the University of Applied Sciences of The Hague (Netherlands). Using a Posttest-Only Control Group Design, participants were assigned to either a control group or a treatment group receiving real-time emotional feedback via custom software. This feedback, based on facial recognition technology, aimed to enhance team performance by providing immediate insights into the team's emotional state.

Over 29 experiments, with 28 yielding analyzable data, participants engaged in a collaborative Mars colony simulation game. The collected data included video and audio recordings, capturing a wide range of emotional and conversational dynamics. We aimed to understand how real-time emotional feedback influences team performance and dynamics during these interactions. By analyzing this multimodal data with machine learning models, we aimed to predict team performance based on emotional and conversational indicators.

Our research contributes to the field of virtual team dynamics and performance in three significant ways. First, we demonstrate how real-time emotional feedback can enhance virtual team performance, providing a novel approach to team collaboration. Second, we extend existing research on the role of emotions in organizational behavior by incorporating multimodal data analysis to capture the nuances of team interactions. Third, we highlight the predictive power of machine learning models in analyzing multimodal data to forecast team performance, showcasing the potential of advanced analytical techniques to develop more effective virtual collaboration tools.

By integrating quantitative analysis and model-based investigation, this study aims to provide a solid foundation for our research questions. The insights gained contribute to the academic understanding of virtual team dynamics and offer practical implications for improving team performance in the digital workplace. Our findings underscore the potential of real-time emotional feedback to optimize virtual team interactions, paving the way for future research and innovation in this area.

3.1. Sample Characteristics

While designed and prepared at MIT, the execution of the experiment spanned multiple universities to amass a larger dataset, enhancing the study's validity and applicability. The study sample consisted of students and faculty from three institutions: the Massachusetts Institute of Technology (MIT) in the USA, Tecnológico de Monterrey in Mexico, and the University of Applied Sciences of The Hague in the Netherlands. This culturally diverse participant pool included 84 individuals with an average age of 23.8 years ($SD = 6.77$), ranging from 18 to 64 years. Approximately two-thirds of the participants were male. Teams were predominantly mixed-gender ($n=18$), with male-only teams constituting a smaller portion ($n=10$); there were no female-only teams. Participants were required to have a working knowledge of English for effective communication in international teams.

3.2. Sampling

The experiments were conducted between November 1, 2023, and January 14, 2024. Participant recruitment was facilitated through various channels, including mailing lists, classroom announcements, and direct contact, leveraging university networks such as the SDM master's program at MIT. Participation was voluntary and uncompensated. Participants were assigned to either the control or treatment groups using a randomized sequence based on their sign-up order. To maintain logistical efficiency, participants were not mixed between sites. Experiments were

conducted on a rolling basis, allowing participants to choose time slots that suited their availability, despite the significant coordination efforts required to assemble groups of three for each session.

The study successfully conducted 29 experiments, with 28 being analyzable. The experiments took place across three locations: 14 in Cambridge, USA; 13 in Monterrey, Mexico; and 3 in The Hague, Netherlands. Despite minor technical issues in some experiments, all were retained for analysis. Exclusion criteria included non-uniform team sizes, leading to the removal of one group that consisted of only two participants. Technical issues necessitated reassigning some groups between control and treatment conditions, but all adjustments were carefully documented to ensure the integrity of the experimental design.

3.3. Procedure

The experiment, leveraging facial recognition software for emotion recognition, received approval from MIT’s Committee on the Use of Humans as Experimental Subjects (COUHES). The study employed a Posttest-Only Control Group Design, as outlined by Campbell and Stanley [55], to evaluate the impact of real-time emotional feedback on virtual team performance. Participants engaged in a collaborative problem-solving task, the Mars Colony simulation game [93], which demands effective communication and coordination. The treatment group was provided with real-time feedback on their emotional states using the Moody software. This system employs cameras to analyze participants’ facial expressions [91], and delivers immediate feedback via a virtual mirroring dashboard [92]. The control group did not receive any feedback.

Having explained the IPO model in the field of virtual teams in section 2.5 and having described our research variables, we contextualized the research design within this framework. Figure 1 integrates the IPO model with the variables relevant to our experiment and provides a schematic representation that illustrates their interplay and importance in our study and is adapted from McGrath (1964) [16].

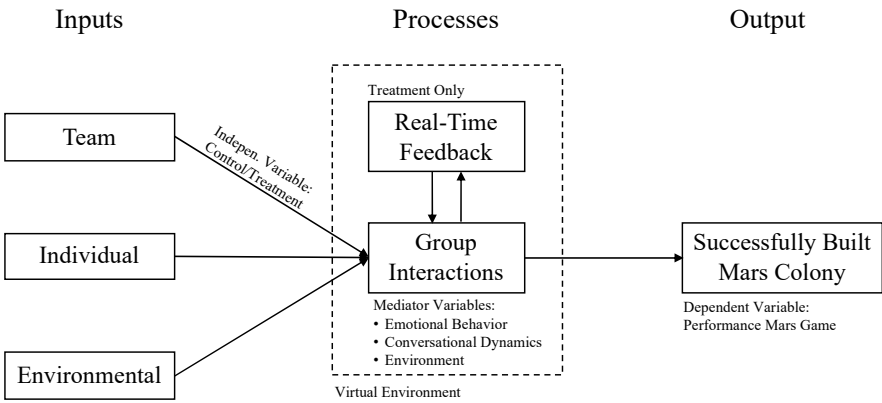


Figure 1. IPO Model for Virtual Teams.

For the initial setup, the Mars Colony game and a real-time emotional feedback dashboard built into Moody [91, 92] (for the treatment group) were arranged on the desktop. Participants were welcomed and instructed to keep their cameras and microphones on throughout the experiment. The facilitator shared the game instructions and allowed five minutes for participants to read them. This reading period was included in the total game duration of 60 minutes. After the initial instructions, the facilitator gave a quick tour of the Mars simulation game and, for the treatment group, explained the Moody dashboard and its metrics.

Participants were given remote control over the desktop, and the facilitator started the Moody and Zoom recordings. The facilitator then muted themselves and turned off their camera to allow participants to focus on the game. The game session lasted for approximately 50 minutes, with the facilitator monitoring progress and intervening only for technical issues. Participants were reminded to finalize their Mars colony configuration five minutes before the end of the session.

After the game, participants completed a post-experiment survey that collected demographic information, assessed emotional intelligence using the "Reading the Mind in the Eyes" test, and gathered data on participants' familiarity with other team members and simulation games. This survey aimed to control for potential confounding variables that could influence the study's outcomes.

The data collected from the Zoom recordings, including video and audio, was used to analyze team interactions, emotions, and communication patterns. The Moody software provided real-time feedback based on facial expressions but was not used for final analysis due to the availability of more advanced emotion analysis tools also developed at the CCI [56, 57]. The Mars simulation game data, including the history of configuration changes, was used to determine team performance.

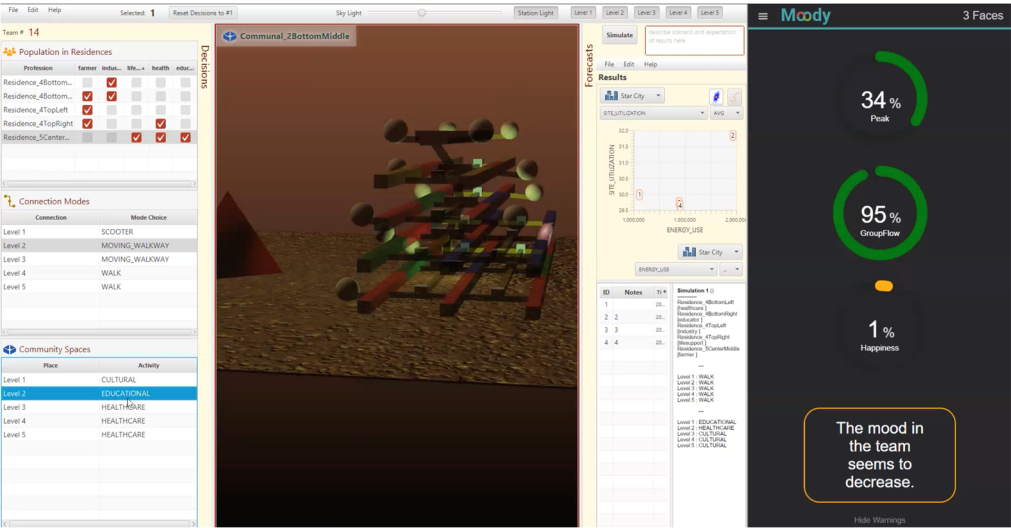


Figure 2. Left side shows the Mars Game, shown to both groups. The right side shows the Moody Dashboard [92] only shown to the treatment group.

3.4. Measures

3.4.1. Dependent Variable

The primary dependent variable in this study is team performance, assessed through the Mars simulation game. This game requires teams to design an optimal Mars colony configuration, balancing "Energy Efficiency" and "Site Utilization." Performance is quantified using the Pareto Score, a metric derived from multi-objective optimization principles that ranks solutions based on their efficiency in achieving both objectives. The Pareto Score evaluates team performance in multi-dimensional optimization problems [58, 59]. It identifies non-dominated solutions that minimize energy use and maximize site utilization, eliminating subjectivity in assessing team effectiveness. Mathematically, for two solutions A and B, A is non-dominated if its energy use is less than or equal to that of B ($Energy(A) \leq Energy(B)$) and its site utilization is greater than or equal to that of B ($Site\ Utilization(A) \geq Site\ Utilization(B)$). Solutions that do not meet these conditions are deemed dominated by those that do. Non-dominated solutions are assigned the highest rank (Rank 0), and the process of exclusion and re-evaluation continues, assigning subsequent ranks to new sets of non-dominated solutions.

Figure 3 presents a visualization that illustrates the hierarchy of Pareto Ranks across the tradespace of possible solutions i.e. all the possible Mars colony configurations and the resulting Energy Use and Site Utilization values. This graphical representation aids in understanding the trade-offs and relative performance of different team site configurations within the multidimensional optimization framework.

To provide a robust measure of team performance, several metrics were derived from the Pareto Ranks. One simple approach would be to just take the average of all Pareto scores. However, this

approach is susceptible to outliers. To mitigate this, performance improvement was calculated as the difference in average ranks between the first and second halves of the task, indicating whether teams improved over time. Another key metric, the regression slope, is obtained by fitting a regression line to the Pareto ranks over time, focusing on values between the 10th and 85th percentiles to minimize outlier effects. A negative slope in this context indicates an improvement in performance. This regression value is further categorized into low, middle, and high values to explore whether such binning of the slope variable offers a more effective approach for the subsequent analysis.

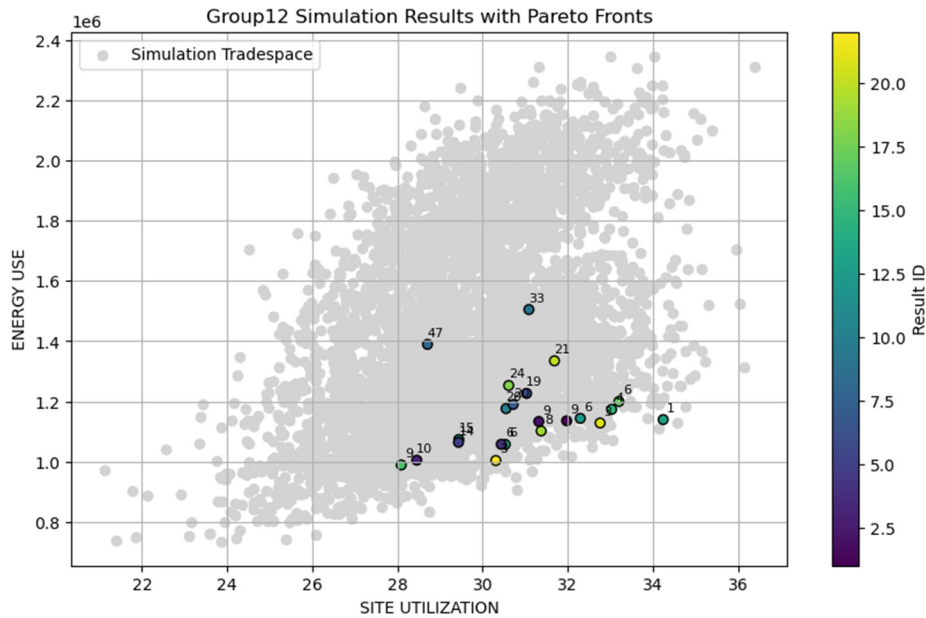


Figure 3. The gray dots represent the space of all possible solutions (tradespace). The colored dots represent the actual simulations of Group12 with the respective Pareto Rank.

3.4.2. Independent Variables

The independent variable in this analysis is whether a team received real-time feedback (treatment group) or not (control group). To measure the effect of feedback on team performance, several mediating variables were collected. These include emotional expressions and communication patterns within the team, derived from facial and audio analysis software. For each metric, the mean, median, standard deviation, max/min value, slope, and percentiles were calculated.

The Facial Analysis System (FAS) [62] analyzed video recordings to detect seven discrete emotions (neutral, surprised, happy, fearful, disgusted, angry, sad), along with 3D head poses and brightness levels. This system provided VAD-values (valence, arousal, dominance) and head motion patterns, contributing to a comprehensive understanding of non-verbal cues and interpersonal dynamics within the team, see Table 1. The Audio Analysis System (AAS) [57] evaluated voice data to derive emotional expressions and communication patterns. This analysis included VAD-values, speaking duration, number of utterances, and interruptions, offering a nuanced view of the communication dynamics within the team, see Table 2. Both systems generated extensive datasets, encompassing key emotional and behavioral statistics.

Table 1. Variables of the FAS.		
Category	Variable	Description
Emotion Metrics	Discrete emotions	Neutral, surprised, happy, fearful, disgusted, angry, sad
	VAD-values	Valence, arousal, dominance
	Emotions Max Count	Maximum number of times a particular emotion is expressed

	Frequency Emotion Changes	Frequency of changes in emotional expressions
Other Non-verbal Cues	Head Velocity	Speed at which a person's head moves
	Brightness	Reflects the lighting conditions of the environment
	Presence	Measures how much time an individual is visible on camera

Table 2. Variables of the AAS.

Category	Variable	Description
Emotional Features	VAD-values	Valence, arousal, dominance
	Speaking Duration	Both absolute and relative speaking times are calculated.
Communication Patterns	Number of Utterances	Quantifies how often a person speaks.
	Number of Interruptions	Calculates how often a speaker interrupts other people in the team.

3.4.3. Control Variable

To ensure the robustness of the analysis, several control variables were included to account for potential confounding factors. Team composition was considered, categorizing the gender composition of the teams as male-only or mixed-gender, with no female-only teams. The level of acquaintance among team members was assessed through the post-experiment survey, capturing whether participants knew each other prior to the experiment. Emotional intelligence was measured using the "Reading the Mind in the Eyes" (RMET) test [60], which evaluates participants' ability to interpret emotions from eye expressions. This inclusion aimed to determine whether the outcomes were influenced by participants' inherent emotional intelligence or by the effectiveness of our intervention [60]. Additionally, the time of day at which the experiment was conducted was recorded, as performance could vary throughout the day due to factors like fatigue or circadian rhythms [61]. The location of the experiment, whether conducted at MIT, Tecnológico de Monterrey, or the University of Applied Sciences of The Hague, was also considered to account for any site-specific effects.

3.5. Data Analysis Approach

3.5.1. Data Preparation

To ensure the robustness of the analysis, both datasets, which were susceptible to outliers due to their nature of collection, were initially prepared for further processing [62, 63]. Gaussian-distributed features were standardized using scikit-learn to ensure they have a mean of zero and a standard deviation of one [64]. Non-Gaussian features underwent robust scaling, which scales the data based on percentiles rather than the mean and standard deviation, making it less sensitive to outliers. Outliers, defined as values with a z-score greater than 3, were replaced with the median value of the respective feature.

Additionally, to facilitate seamless integration of data from facial analysis and audio analysis, a unique mapping was created to match the IDs between the two output files. After this process, columns deemed irrelevant to the analysis, such as those relating to time or IDs, were removed. For the regression of Pareto Scores, only values between the 10th and 85th percentile were considered, aiming to mitigate the influence of outliers in the subsequent analysis.

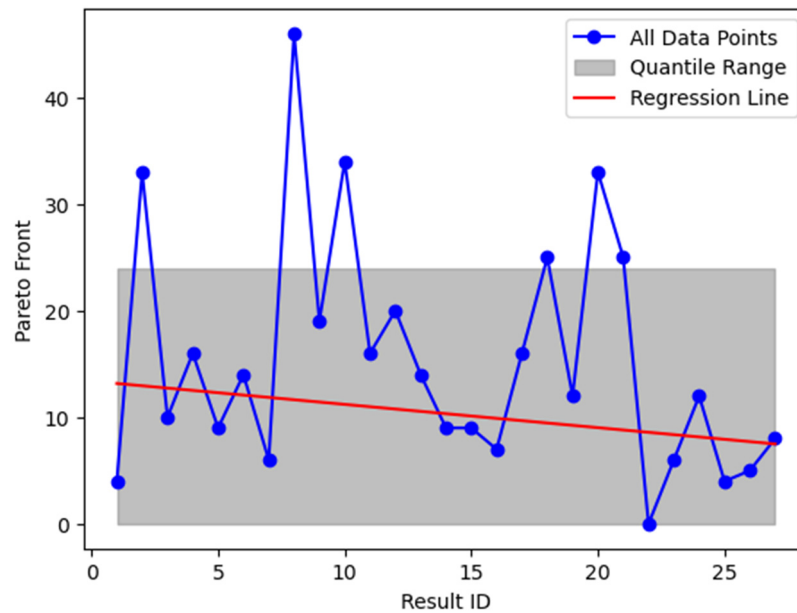


Figure 4. Visualization of the Performance Slope Calculation (Example).

3.5.2. Descriptive Statistics

Before delving into the multi-level mixed-effects linear regression analysis, we first examined the relationships between the predictor variables. We created a correlation matrix to assess multicollinearity, using both Pearson correlation coefficients and Variance Inflation Factors (VIFs) [65]. Heteroscedasticity was also examined using the Breusch-Pagan test [66, 67]. To understand the most suitable variables for our analysis, Pearson correlation was calculated between each predictor variable and the various performance metrics.

In addition, we utilized t-tests and Mann-Whitney U tests to identify statistically significant differences between the control and treatment groups [68, 69]. To quantify the magnitude of these differences, we calculated effect sizes using Cohen's *d* for normally distributed data and the rank biserial correlation for non-normally distributed data [70, 71].

3.5.3. Multi-Level Mixed Effects Linear Regression

Recognizing the hierarchical structure of our data (individuals nested within teams, and teams nested within groups), we employed a mixed-effects model for our regression analysis [72]. This allowed us to account for both fixed effects (e.g., real-time emotional feedback) and random effects, which capture the variations occurring within teams and groups [73].

A null model with only random intercepts was used as a baseline for assessing the Intraclass Correlation Coefficient (ICC), providing insights into the proportion of variance explained by the grouping structure. In our main models, we included several control variables such as team composition, prior acquaintance among team members, and time slots to account for potential confounding effects. The performance of the models was evaluated using the Akaike Information Criterion (AIC), a measure of relative model quality, as well as marginal and conditional R^2 , which quantify the variance explained by the fixed effects and the entire model respectively [74-76].

3.5.4. Feature selection

Feature selection was performed to improve the model's ability to generalize to new data and prevent overfitting [77]. Subsets of features were generated based on the insights gained from correlation analysis and VIF calculations, and then utilized for further model development. For tree-based models such as Random Forests and XGBoost, an inherent feature selection mechanism is

embedded in the training process. For the Support Vector Machine (SVM), a more refined dataset, produced after the VIF calculation, was employed for model training.

3.5.5. Training of Machine Learning Models

The machine learning models, including Support Vector Machine (SVM), Random Forests, XGBoost, and NGBoost, were trained to predict team performance based on the various features extracted from the dataset [78-83]. Additionally, K-Means clustering was employed to identify potential hidden patterns or groupings within the data [84]. Due to the continuous nature of our target variable, which represents team performance, this was framed as a regression problem. Hyperparameter tuning for each model was conducted using both randomized and grid search techniques. An 80/20 train-test split strategy was employed to ensure robust model evaluation.

The models were evaluated based on several key performance metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 . The MAE and RMSE provide measures of the average magnitude of prediction errors, while the R^2 represents the proportion of variance in the target variable explained by the model [85].

3.5.6. SHAP-Values for Feature Interpretability

SHAP values were calculated to assess the importance of each feature in the model and to understand their contributions to the final prediction [86-88]. SHAP values provide a unified approach to explain the output of any machine learning model. In our analysis, SHAP values were used to rank features by their importance, allowing us to identify the most influential factors in determining team performance.

4. Results

4.1. Correlation Analysis, Feature Elimination, and Heteroscedasticity

An initial correlation analysis was performed on 186 features derived from the facial and audio analysis software. Given the limited sample size ($n = 84$), a correlation matrix with a threshold of 0.7 was used to identify and remove highly correlated features, reducing the feature set to 50. To address the potential issue of multicollinearity among the remaining features, the Variance Inflation Factor (VIF) analysis was employed, setting a threshold of 10. Outliers, identified with z-scores of six or more, were replaced with the median, further trimming the features to 31 [89]. This mitigated multicollinearity and refined the feature set for subsequent analyses. The Breusch-Pagan test revealed no evidence of heteroscedasticity (Lagrange Multiplier $p = .240$, F-statistic $p = .171$).

4.2. Correlation Analysis with Dependent Variable

Correlation analysis, using Pearson coefficients with a significance threshold of $p \leq .10$, was conducted to explore relationships between the dependent variables ("Performance Slope," "Performance Slope Binned," and "Pareto Rank Difference") and the remaining features. "Performance Slope" emerged as the most suitable dependent variable due to its numerous and strong correlations with features like vocal arousal (Vocal Arousal (Slope), $r = .34$, $p = .0016$), emotional expression (Neutral (Min), $r = .31$, $p = .0039$), and engagement (Happy (Slope), $r = .29$, $p = .0071$). Therefore, we decided to focus exclusively on the "Performance Slope" variable for subsequent analyses.

4.3. Differences Between Control and Treatment Group

T-tests and their non-parametric equivalents were employed to evaluate differences between our control and treatment groups, see Table 3. The tests revealed significant differences in several key variables, suggestive of the impact of our intervention. To quantify the magnitude of these differences, effect sizes were calculated. For the t-tests, Cohen's d was used [70], and for the Mann-

Whitney U tests, where the data were not normally distributed, the rank biserial correlation (rrb) was utilized [71].

Table 3. Differences Across Control and Treatment Groups.

Feature	Test Used	P-Value	Effect Size
Brightness (Slope)	Mann-Whitney U	.0046	-.17***
Happy (Std)	Mann-Whitney U	.0063	.19***
Performance Slope	Mann-Whitney U	.0159	.17**
Abs. Utterances (Std)	T-test	.0165	-.53**
Happy (Median)	Mann-Whitney U	.0198	.16**
Facial Arousal (Max)	Mann-Whitney U	.0634	.13*
Surprise (Slope)	Mann-Whitney U	.0788	-.10*
Facial Valence (Max)	Mann-Whitney U	.0900	.12*
Abs. Interruptions (Min)	Mann-Whitney U	.0958	.11*

Table displays significant results with $p \leq .10$. Positive effect size indicates that the treatment group has a higher score on the measured outcome than the control group. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The variable “Performance Slope,” indicative of team performance, evidenced a significant difference between control and treatment groups, signifying a measurable though small impact of our interventions ($rrb = .17$). Effect sizes varied for other variables. The largest effect size was observed for Absolute Utterances (Std) ($d = -.53$), suggesting more uniform participation within the treatment group compared to the control group, indicating that the treatment may have encouraged more equitable participation.

Emotionally, the treatment group exhibited higher variability in happiness (Happy (Std); $rrb = .19$), suggesting that our intervention introduced a broader range of emotional responses among team members. Additionally, the modest rise in median happiness ($rrb = .16$) indicates a general uplift in mood among team members, suggesting that the average emotional state was positively affected by the intervention.

The significant difference in brightness slope ($rrb = -.17$) indicates that the treatment group experienced environments with less variation in illumination. Lower illumination is shown to affect mood and focus [90].

While variables like Facial Arousal (Max), Surprise (Slope), Facial Valence (Max), and Absolute Interruptions (Min) did not achieve conventional statistical significance, their effect sizes and p-values indicate relevant trends ($rrb = .13, -.10, .12$, and $.11$, respectively). These findings suggest that the intervention influenced performance, emotional, and communicative dynamics within teams.

4.4. Multi-Level Mixed Effects Regression

A multi-level mixed effects regression model was employed to account for the nested data structure (individuals within teams, teams within groups). With 84 data entries spread across 28 teams, we opted to group our model by the variable Group to categorize teams into treatment and control groups. Simultaneously, we modeled the individual teams as a random effect within our null model. This approach better reflects the nested structure of the data while accounting for inherent variability at the team level (table 4).

The general structure of our approach involved creating a null model to establish a baseline, adding control variables to account for confounding factors, and then constructing a full model that integrates all relevant predictors. This approach aligns with the general structure of hierarchical linear models [73].

The null model ($AIC = 252.94$) revealed negligible variance at the group level ($ICC = 0.0086$). The Marginal R^2 and Conditional R^2 for the null model were 0.0099 and 0.0184, respectively, showing the low variance explained by the group-level random effects alone. The null model indicated that neither the baseline level of Slope nor the variation across team numbers significantly contribute to the model’s predictions, as shown by the non-significant intercept and the random effects of team

numbers. There is a small variance component for the grouping variable, suggesting minimal variability in performance due to differences between control and treatment groups.

The inclusion of control variables (gender composition, experiment time, emotional intelligence, prior acquaintance) improved the model's explanatory power (average AIC = 251.65, marginal R2 = 0.1431, conditional R2 = 0.1655), highlighting the importance of gender diversity (Gender Composition, $p < .05$) and pre-existing social ties (Acquaintance, $p < .05$) for team performance.

The full model, incorporating all 31 features obtained after the VIF calculation, further enhanced the model's explanatory power (average marginal R2 = 0.6483, conditional R2 = 0.6604). The AIC also increased to 341.31, which, despite being higher than the control model, is not unexpected due to the introduction of a high number of parameters. Notably, the model revealed a nuanced relationship between interruptions and performance, where absolute interruptions negatively impacted performance (Abs. Interruptions (Mean), $p < .01$), but a positive trend in relative interruptions over time was associated with enhanced performance (Rel. Interruptions (Slope), $p < .05$). Additionally, increased communication volume (slope of absolute utterances) correlated with better performance (Abs. Utterances (Slope), $p < .1$), while excessive brightness negatively affected performance (Brightness (Median), Brightness (Std), $p < .1$). The model also highlighted the detrimental effect of high fear expressions (Fear Count (Max), $p < .1$) and the positive influence of gender diversity (Gender Composition, $p < .05$). Striking is the significant negative coefficient for the Fear Count (Max) ($\beta = -94.161$, $p = .053$), which indicates a strong negative influence of the maximum number of fear expressions on team performance. This could suggest that high levels of fear or stress expressions within a team affect its effectiveness, emphasizing the importance of the emotional climate in team interactions.

Table 4. Overview of Models of Multi-Level Mixed Effects Regression.

Metrics / Features	Null Model	Control Model	Full Model
AIC	252.94	251.65	341.31
Marginal R2	0.0099	0.1431	0.6483
Conditional R2	0.0184	0.1655	0.6604
Intercept	0.175	0.839	1.100
Team Random Effect	-0.013	-0.014	0.003
Gender Composition		0.262*	0.396**
Experiment Time		-0.189	-0.254
RMET Test		-0.488	-1.365
Acquaintance		-0.618**	-0.283
Happy (Std)			-0.172
Fear (Min)			0.006
Neutral (Std)			0.206
Neutral (Min)			-0.296*
Facial Valence (Max)			0.014
Facial Arousal (Max)			0.005
Facial Arousal (Min)			0.216
Facial Dominance (Min)			-0.029
Brightness (Median)			-0.316*
Brightness (Std)			-0.489*
Brightness (Min)			0.328
Brightness (Slope)			-0.081
Velocity (Median)			-0.142
Velocity (Max)			0.090
Velocity (Min)			0.181
Velocity (Slope)			-0.002
Fear Count (Max)			-94.161*

Vocal Arousal (Mean)	0.042
Vocal Arousal (Slope)	0.269
Vocal Arousal (Std)	0.050
Vocal Valence (Slope)	0.022
Abs. Interruptions (Mean)	0.337
Abs. Interruptions (Slope)	-0.320***
Abs. Interruptions (Min)	-0.105
Rel. Interruptions (Mean)	-0.028
Rel. Interruptions (Slope)	0.306**
Rel. Interruptions (Max)	-0.025
Abs. Utterances (Slope)	0.213*
Abs. Utterances (Std)	-0.192
Rel. Utterances (Std)	0.060
Abs. Speak Duration (Std)	0.227

Note. *p < 0.1; **p < 0.05; ***p < 0.01.

4.5. Prediction with Machine Learning Models

4.5.1. Model Evaluation

Hyperparameter optimization was conducted for Random Forest, XGBoost, NGBoost, and SVM models (table 5). XGBoost outperformed other models, demonstrating the lowest MAE (0.3522) and MSE (0.3123) and the highest R2 (0.6762), indicating superior predictive accuracy. SVM performed the worst, with the highest RMSE (1.0055) and a negative R2 (-0.0483).

Table 5. Differences Across Control and Treatment Groups.

Model	MSE	MAE	RMSE	R ²
Random Forest	0.6894	0.6289	0.8303	0.2853
XGBoost	0.3123	0.3522	0.5588	0.6762
NGBoost	0.4144	0.4888	0.6437	0.5704
SVM	1.0111	0.5549	1.0055	-0.0483

4.5.2. Unsupervised Learning with K-Means Clustering

The data, already standardized, required no further processing before K-Means clustering. Using the elbow method, the optimal number of clusters was determined to be four (k=4), as the graph of Within-Cluster-Sum-of-Squares (WCSS) flattened after this point. Cluster centroids were computed to analyze average feature values and the distribution of treatment and control group participants. Four distinct clusters emerged, each with unique characteristics in emotional expressions and engagement metrics (table 6).

Cluster 0 ("Reserved and Steady") exhibited consistent emotional responses and moderate engagement. Cluster 1 ("Expressive and Dynamic") displayed greater emotional variability and higher engagement. Cluster 2 ("Neutral and Bright") showed neutral emotional behavior and significant environmental brightness. Cluster 3 ("Low Engagement") was characterized by lower brightness and reduced emotional and interaction dynamics. Cluster 1 had the highest mean (0.4652) and median (0.1445) "Performance Slope," suggesting a link between expressiveness and dynamic engagement with performance.

Table 6. Differences Across Control and Treatment Groups.

Cluster	Count	Count Control	Count Treatment	Mean Performance	Median Performance
0	29	14	15	0.0128	0.0942

1	21	6	15	0.4652	0.1445
2	19	13	6	-0.5169	-0.5585
3	15	9	6	-0.1256	-0.1201

4.5.3. Feature Interpretation with SHAP Values

SHAP values were analyzed for NGBoost, Random Forest, and XGBoost models to understand feature importance. The SVM model was omitted because of its low performance. NGBoost placed significant emphasis on Velocity (Min), Happy (Std), and Facial Valence (Median), indicating that this model values both physical movement and emotional variability. This suggests NGBoost's sensitivity to changes in emotional expressions and their potential impact on outcomes.

Random Forest and XGBoost, on the other hand, prioritized Vocal Arousal (Slope), Velocity (Min), Neutral (Std), and Brightness (Slope), albeit with different weights. XGBoost particularly emphasized Vocal Arousal (Slope), highlighting its sensitivity to changes in vocal pitch over time, which could signal team enthusiasm, stress, or agreement. Random Forest additionally emphasized Fear (Min), while XGBoost also considered Sad (Median) and Happy (Slope), reflecting a broader approach to capturing emotional health.

A commonality across all models was the importance of Velocity (Min), underscoring its overall relevance, likely capturing crucial aspects of physical activity or movement pertinent to team performance. However, the models diverged in their prioritization of other features. NGBoost's focus on Happy (Std) contrasts with Random Forest and XGBoost's focus on Vocal Arousal (Slope), illustrating NGBoost's focus on emotional variability while the other models emphasize the temporal dynamics of vocal expressions.

5. Discussion

This study aimed to investigate the influence of real-time emotional feedback on virtual team performance. Data were collected over two months, with teams participating in a simulation game across various locations. Statistical analyses revealed marginally significant differences in performance between the control and the group receiving real-time emotion feedback (treatment), alongside notable variations in emotional and conversational behaviors, suggesting an impact on social dynamics. Multimodal analysis employing machine learning algorithms indicated the ability to capture substantial variance in team performance.

5.1. Predictability of Team Performance

The findings support the hypothesis that multimodal data can accurately predict virtual team performance, aligning with previous research [48, 81, 84]. However, the small dataset and specific experimental setup necessitate further studies to generalize these findings across different contexts and larger, more diverse samples.

Contrary to initial hypotheses, results reveal a nuanced impact of emotional states and physical dynamics on team performance. While positive emotions play a role, the standard deviation of happiness, changes in vocal arousal, and minimal movement emerged as critical predictors. This underscores the complex role of emotional dynamics and physical engagement in team effectiveness, suggesting that both positive states and arousal changes, alongside active participation, contribute to team performance.

Conversational variables, while significant in regression analysis, were less prominent in machine learning models, challenging the hypothesis that they are primary performance predictors. This suggests that emotional and physical engagement may play a more dominant role, diverging from literature emphasizing equitable participation [15]. The low importance of conversational behavior in our models might be attributed to the complex nature of group dynamics, where the impact of such behaviors can vary significantly depending on context [18]. While conversational dynamics are traditionally valued, their predictive power may be overshadowed in environments where emotional intelligence and physical engagement are more pivotal [23].

5.2. Effect of Treatment on Team Performance and Dynamics

Our third hypothesis stated that teams receiving the treatment would show improved performance scores. The analysis found a significant, albeit marginal, impact on performance, thus validating our hypothesis. However, regression analysis revealed negligible variance between control and treatment groups, challenging the robustness of this hypothesis. Data collection inaccuracies further caution the generalizability of our findings. This aligns with Tausczik and Pennebaker [89], who found real-time language feedback effective only in well-coordinated teams.

For Hypothesis 4, which examined the effect of real-time emotional feedback on team emotional regulation, results were nuanced. Contrary to the hypothesis predicting reduced variance, we observed increased variability in happiness, indicating that feedback broadened emotional responses. This was accompanied by slight increases in positive expressions and moderate surprise, suggesting a more positive emotional climate. These findings challenge Hypothesis 4, showing that real-time feedback enriches emotional dynamics rather than streamlining them [50].

Hypothesis 5, suggesting a broader emotional spectrum in the treatment group, was only partially supported. Higher levels of happiness and facial valence were observed, but other emotions showed no significant differences. This indicates that the intervention partially broadened the emotional range and elevated team mood. The lack of a pronounced emotional spectrum suggests further analysis is needed. Existing research highlights the importance of feedback in enhancing group emotional intelligence but often does not establish a clear link between feedback and a broader range of emotional expressions [54].

In conversation patterns, significant differences were observed in communication variability and fewer interruptions, highlighting feedback's potential to balance speaking turns and promote constructive interruptions. Real-time feedback modestly influenced the emotional landscape, resulting in increased emotional variability and uplifted mood. This suggests the intervention facilitated more nuanced emotional engagement, contributing to a richer and more positive team environment, and demonstrating the potential of real-time feedback to enhance emotional adaptability and diversity in teams.

Hypothesis	Accepted/Rejected	Reason
Hypothesis 1	✓	High predictive accuracy with XGBoost ($R^2 = 0.6762$).
Hypothesis 2a	✓	Positive emotional states correlated with better performance ($r = .29$, $p = .0071$).
Hypothesis 2b	✓	More uniform participation observed in the treatment group ($d = -.53$).
Hypothesis 3	✓	Significant difference in Performance Slope ($p = .0159$, $rrb = .17$).
Hypothesis 4	✗	Increased variability in happiness suggests feedback broadened emotional responses.
Hypothesis 5	✗	No significant increase in emotional spectrum metrics other than happiness.

Figure 5. Acceptance and rejection of proposed hypothesis.

5.3. Behavioral Archetypes in Virtual Team Dynamics

Unsupervised clustering analysis revealed four distinct team behavioral archetypes, each with different interaction and performance patterns. Teams with consistent emotions and moderate engagement ("Reserved and Steady") may prioritize efficiency, while those with greater emotional variability and higher engagement ("Expressive and Dynamic") may foster creativity. Conversely, teams with limited emotional expression and engagement ("Low Engagement") struggle with collaboration. This underscores the importance of emotional and engagement metrics in virtual team

performance and suggests that tailoring interventions to specific team archetypes could be more effective than a one-size-fits-all approach.

5.4. Limitations and Future Research

Our study acknowledges certain limitations, particularly the relatively small dataset, which may affect the robustness of our conclusions due to the high number of features compared to data points, known as the “curse of dimensionality” [90]. While our dataset is sufficient for initial conclusions, expanding it would enhance the depth and validity of our analysis, mitigate individual group variability, and improve the generalizability of our findings. Future research should focus on increasing the dataset size and refining the treatment approach. The general nature of the information provided in our experiment suggests that more tailored and actionable feedback, possibly utilizing advanced large language models (LLMs), could better prompt behavioral modifications in participants.

Another factor of uncertainty is our metric for measuring team performance. Despite efforts to create an objective assessment, the inherent variance in the simulation game outputs diluted the performance evaluation. We attempted to counteract this by examining performance trends over time and using a specific quantile range to reduce outlier impact. However, this approach does not entirely eliminate potential biases and may overlook instances of exceptional performance or significant failure. Additionally, regression analysis assumes consistency across iterations, which might miss complex, nonlinear interactions influencing team performance.

Our software for analyzing emotions and communication patterns also poses limitations. Despite advancements in facial and speech recognition models, inaccuracies persist, especially given our use of slightly outdated software and the real-world experimental conditions, which introduced challenges like partial visibility of participants' faces. These factors likely contributed to inaccuracies in emotion recognition.

Feature selection was another challenge, as examining a wide range of metrics required working with many features relative to our dataset size. While this aimed to capture different aspects of team dynamics, it was necessary to exclude some features to improve interpretability, potentially losing explanatory power.

Our results showed significant differences between the experimental and control groups, but the extent of these differences was small and uncertain, warranting cautious interpretation. Our experimental design did not account for the potential impact of feedback presence alone. Introducing a third group receiving non-specific randomized feedback in future studies could help determine whether observed effects stem from targeted feedback or the feedback mechanism itself. This approach would clarify the direct effects of emotional feedback versus general monitoring effects and investigate long-term impacts on team cohesion and trust.

Future research should also explore how feedback is perceived, focusing on technology acceptance to enhance the adoption of real-time feedback technologies. This involves examining usability, perceived usefulness, and resistance to monitoring. By gathering user feedback and applying user experience research methodologies, feedback mechanisms can be refined to better meet team needs. Investigating different feedback content, timing, and methods, and comparing feedback modalities (e.g., visual, auditory, textual) will be crucial for understanding their impact on virtual team dynamics. Additionally, exploring cultural differences in feedback reception and effectiveness in global virtual teams is essential to ensure the universal applicability of these tools. Integrating biometric data could add depth to team dynamics analysis, and developing automated algorithms for early conflict detection and resolution could serve as proactive measures to maintain team cohesion.

5.5. Implications

The observation of the different effects of real-time emotional feedback on team performance adds empirical evidence to the discourse on emotional intelligence and team dynamics. This aligns with Druskat and Wolff's notion that a group's emotional intelligence is a catalyst for effective team

interaction [54], as well as Handke et al.'s findings on the positive impact of emotional awareness tools [50]. Notably, the identification of physical engagement indicators, such as head movement speed, as key predictors of team performance broadens our understanding of team dynamics beyond emotional states to include non-verbal cues. This suggests a complex interplay of factors influencing virtual team performance, highlighting the potential of real-time feedback mechanisms to enhance both the emotional and physical aspects of virtual teamwork.

The practical implications of this research are significant for developing virtual collaboration tools. Integrating emotional insights could help teams regulate emotions and foster positive environments, potentially boosting performance. However, ethical considerations regarding privacy and consent are crucial. Future collaboration platforms could incorporate emotional analysis and feedback while prioritizing user privacy and ethical guidelines to balance useful monitoring with emotional surveillance.

To enhance virtual teamwork, it is essential to design inclusive feedback mechanisms accommodating diverse emotional expressions and communication styles across cultures. This approach enriches team dynamics and aligns with evidence supporting the positive impacts of emotional intelligence and open communication in teamwork [14, 15]. Inclusivity can help leaders cultivate policies that leverage technology to strengthen team cohesion and performance, through training programs focused on improving emotional intelligence, communication strategies, and conversational dynamics management. These skills are valuable for navigating virtual teamwork and face-to-face interactions.

Identifying emotional states, conversational patterns, and physical dynamics as important predictors highlights the complex structure of virtual team dynamics. This complements theoretical frameworks by Barsade and Knight [46] and Meneghel et al. [47], emphasizing the importance of emotional coherence and conversational balance in team performance. These findings call for a deeper examination of the interplay of emotional and conversational factors in different virtual team settings.

The effective application of machine learning models, such as XGBoost, to uncover insights into team performance reflects the growing relevance of computational analysis in understanding team interactions. This aligns with the broader shift towards data-driven approaches in social sciences [48, 80, 81, 84]. Future research should explore these models in diverse team contexts and leverage the predictive power of multimodal data to develop intervention tools that offer nuanced feedback to teams, potentially identifying and addressing emerging issues proactively.

6. Conclusions

In this exploration of virtual team dynamics in a digitally evolving work environment, key factors that contribute to team success have been identified, emphasizing the nuanced role of emotional and conversational dynamics. This experimental study showed that the treatment – real-time emotional feedback on the team level – evidenced a marginal yet statistically significant improvement in team performance. The subtleties of their influence on emotional and conversational behavior underscore the complexity of social dynamics in virtual teams. Teams characterized by emotional diversity, stability, and positive valence, as well as more balanced conversational dynamics are better positioned for success, highlighting the importance of emotion management in optimizing team performance. By integrating multimodal data, a significant amount of variance in team performance could be captured, highlighting the potential of computational approaches to improve our understanding of team dynamics. Our results should be seen as a contribution to a better understanding of complex team interactions in digital environments, requiring further research to validate and extend these preliminary results. By understanding the complex interplay of factors influencing virtual team dynamics, we can pave the way for more effective collaboration and communication in the digital age.

Author Contributions: Conceptualization, PG and NS; methodology PG and NS; software, NS; formal analysis, NS; investigation, NS and IV; resources, PG and IV; data curation, NS; writing—original draft preparation, NS; writing—review and editing, PG, NS and IV; supervision, PG and IV; project administration, IV and NS; All authors have read and agreed to the published version of the manuscript.

Funding: Nicklas Schneider was funded by the Federal Ministry of Education and Research (BMBF). This work was supported by a fellowship within the IFI programme of the German Academic Exchange Service (DAAD).

Institutional Review Board Statement: This study was approved by MIT COUHES under IRB1701817083 dated 19 January 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

Conflicts of Interest: There is no conflict of interest.

References

1. Allen, J. A., Lehmann-Willenbrock, N., & Rogelberg, S. G. (2015). *The Cambridge Handbook of Meeting Science* (J. A. Allen, N. Lehmann-Willenbrock, & S. G. Rogelberg, Eds.). Cambridge University Press. <https://doi.org/10.1017/CBO9781107589735>
2. Kniffin, K. M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., Bamberger, P., Bapuji, H., Bhawe, D. P., & Choi, V. K. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *American psychologist*, 76(1), 63. <https://doi.org/10.1037/amp0000716>
3. Bailey, D. E., & Kurland, N. B. (2002). A review of telework research: Findings, new directions, and lessons for the study of modern work. *Journal of Organizational Behavior*, 23(4), 383-400. <https://doi.org/10.1002/job.144>
4. Kanse, L., Stephenson, E. K., Klonek, F. E., & Wee, S. (2023). Interdependence in Virtual Teams—A Double-Edged Sword? *Small Group Research*, 10464964231206129. <https://doi.org/10.1177/10464964231206129>
5. Ferrazzi, K. Getting Virtual Teams Right. Available online: <https://hbr.org/2014/12/getting-virtual-teams-right> (accessed on 26 May 2024).
6. Bradbury, T. N., & Fincham, F. D. (1990). Attributions in marriage: review and critique. *Psychological bulletin*, 107(1), 3. <https://doi.org/10.1037/0033-2909.107.1.3>
7. Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-Based Intuition and Decision Making in Organizations. *Journal of Management*, 36(4), 941-973. <https://doi.org/10.1177/0149206309350084>
8. Wespi, R., Birrenbach, T., Schaubert, S. K., Manser, T., Sauter, T. C., & Kämmer, J. E. (2023). Exploring objective measures for assessing team performance in healthcare: an interview study. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1046498>
9. Levi, D., & Askay, D. A. (2020). *Group dynamics for teams*. Sage Publications.
10. Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “Big Five” in Teamwork? *Small Group Research*, 36(5), 555-599. <https://doi.org/10.1177/1046496405277134>
11. Edmondson, A. C., & Schein, E. H. (2012). *Teaming: How Organizations Learn, Innovate, and Compete in the Knowledge Economy*. Wiley-Blackwell.
12. Cohen, P. R., & Levesque, H. J. (1991). Teamwork. *Noûs*, 25(4), 487-512. <https://doi.org/10.2307/2216075>
13. Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review*, 26(3), 356-376. <https://doi.org/10.2307/259182>
14. Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3), 77-124. <https://doi.org/10.1111/j.1529-1006.2006.00030.x>
15. Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686-688. <https://doi.org/10.1126/science.1193147>
16. McGrath, J. E. (1964). *Social Psychology: A Brief Introduction*. Holt, Rinehart and Winston.
17. Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in Organizations: From Input-Process-Output Models to IMOI Models. *Annual Review of Psychology*, 56(1), 517-543. <https://doi.org/10.1146/annurev.psych.56.091103.070250>
18. Mathieu, J. E., Marks, M. A., & Zaccaro, S. J. (2002). Multiteam systems. In *Handbook of industrial, work and organizational psychology*, Volume 2: Organizational psychology (pp. 289-313). Sage Publications, Inc.
19. Jacks, T. Research on Remote Work in the Era of COVID-19. *Journal of Global Information Technology Management*, 2021, 24(2), 93-97. <https://doi.org/10.1080/1097198X.2021.1914500>
20. Cagiltay, K., Bichelmeyer, B., & Kaplan Akilli, G. (2015). Working with multicultural virtual teams: critical factors for facilitation, satisfaction and success. *Smart learning environments*, 2, 1-16. <https://doi.org/10.1186/s40561-015-0028-5>

21. Gibson, C. B., & Cohen, S. G. (2003). Virtual teams that work: Creating conditions for virtual team effectiveness. John Wiley & Sons.
22. Karl, K. A., Peluchette, J. V., & Aghakhani, N. (2022). Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. *Small Group Research*, 53(3), 343-365. <https://doi.org/10.1177/10464964221091968>
23. Standaert, W., Muylle, S., & Basu, A. (2021). How shall we meet? Understanding the importance of meeting mode capabilities for different meeting objectives. *Information & Management*, 58(1), 103393. <https://doi.org/10.1016/j.im.2020.103393>
24. Dulebohn, J., & Hoch, J. (2017). Virtual teams in organizations. *Human resource management review*, 27(4), 569-574. <https://doi.org/10.1016/j.hrmr.2016.12.004>
25. Gaudes, A., Hamilton-Bogart, B., Marsh, S., & Robinson, H. (2007). A framework for constructing effective virtual teams. *The Journal of E-working*, 1(2), 83-97.
26. Friedrich, R., Friedrich, R., & Fetzter. (2017). Virtual Team Maturity Model. Springer.
27. Bickle, J. T., Hirudayaraj, M., & Doyle, A. (2019). Social Presence Theory: Relevance for HRD/VHRD Research and Practice. *Advances in Developing Human Resources*, 21(3), 383-399. <https://doi.org/10.1177/1523422319851477>
28. Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, 32(5), 554-571. <https://doi.org/10.1287/mnsc.32.5.554>
29. Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE transactions on professional communication*, 48(2), 117-130. <https://doi.org/10.1109/TPC.2005.849649>
30. Anderson, N. R., & West, M. A. (1998). Measuring climate for work group innovation: development and validation of the team climate inventory. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 19(3), 235-258. [https://doi.org/10.1002/\(SICI\)1099-1379\(199805\)19:3<235::AID-JOB837>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1099-1379(199805)19:3<235::AID-JOB837>3.0.CO;2-C)
31. Battles, J., & King, H. B. (2010). TeamSTEPPS® teamwork perceptions questionnaire manual. Washington, DC: American Institutes for Research.
32. Goodman, P. S., Ravlin, E., & Schminke, M. (1987). Understanding groups in organizations. *Research in organizational behavior*.
33. Rosen, M. A., Salas, E., Wilson, K. A., King, H. B., Salisbury, M., Augenstein, J. S., Robinson, D. W., & Birnbach, D. J. (2008). Measuring team performance in simulation-based training: adopting best practices for healthcare. *Simulation in Healthcare*, 3(1), 33-41. <https://doi.org/10.1097/SIH.0b013e3181626276>
34. DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of applied psychology*, 95(1), 32. <https://doi.org/10.1037/a0017328>
35. Sun, L., Gloor, P., Stein, M., Eirich, J., & Wen, Q. (2020). "No Pain No Gain": Predicting Creativity Through Body Signals. In (pp. 3-15). https://doi.org/10.1007/978-3-030-48993-9_1
36. Zeyda, M., Stracke, S., Knipfer, K., & Gloor, P. A. (2023). Your body tells more than words – predicting perceived meeting productivity through body signals. *European Journal of Work and Organizational Psychology*, 1-17. <https://doi.org/10.1080/1359432X.2022.2162881>
37. Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of applied psychology*, 92(3), 595. <https://doi.org/10.1037/0021-9010.92.3.595>
38. Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. *Small Group Research*, 43(2), 130-158. <https://doi.org/10.1177/1046496411429599>
39. Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36(4), 876-910. <https://doi.org/10.1177/0149206309356804>
40. Costa, A. C. (2003). Work team trust and effectiveness. *Personnel review*, 32(5), 605-622. <https://doi.org/10.1108/00483480310488360>
41. Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological bulletin*, 90(1), 125-152. <https://doi.org/10.1037/0033-2909.90.1.125>
42. Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative science quarterly*, 47(4), 644-675. <https://doi.org/10.2307/3094912>
43. Frijda, N. H. (1986). The emotions. Cambridge University Press. <https://doi.org/10.1017/CBO9780511805856>
44. Parkinson, B., Totterdell, P., Briner, R. B., & Reynolds, S. (1996). Changing moods: The psychology of mood and mood regulation. Longman London. <https://doi.org/10.4324/9781315003161>
45. Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161. <https://doi.org/10.1037/h0077714>
46. Barsade, S. G., & Knight, A. P. (2015). Group Affect. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 21-46. <https://doi.org/10.1146/annurev-orgpsych-032414-111316>
47. Meneghel, I., Salanova, M., & Martínez, I. (2016). Feeling Good Makes Us Stronger: How Team Resilience Mediates the Effect of Positive Emotions on Team Performance. *Journal of Happiness Studies*, 17, 239-255. <https://doi.org/10.1007/s10902-014-9592-6>

48. Avci, U., & Aran, O. (2016). Predicting the Performance in Decision-Making Tasks: From Individual Cues to Group Interaction. *IEEE Transactions on Multimedia*, 18(4), 643-658. <https://doi.org/10.1109/TMM.2016.2521348>
49. Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3), 77-124. <https://doi.org/10.1111/j.1529-1006.2006.00030.x>
50. Handke, L., Klonek, F., O'Neill, T. A., & Kerschreiter, R. (2022). Unpacking the role of feedback in virtual team effectiveness. *Small Group Research*, 53(1), 41-87. <https://doi.org/10.1177/10464964211067949>
51. Gabelica, C., Van den Bossche, P., Segers, M., & Gijssels, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, 7(2), 123-144. <https://doi.org/10.1016/j.edurev.2011.11.003>
52. London, M., & Sessa, V. I. (2006). Group feedback for continuous learning. *Human Resource Development Review*, 5(3), 303-329. <https://doi.org/10.1177/1534484306290226>
53. DeShon, R. P., Kozlowski, S. W., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of applied psychology*, 89(6), 1035. <https://doi.org/10.1037/0021-9010.89.6.1035>
54. Druskat, V. U., & Wolff, S. B. (2001). Building the emotional intelligence of groups. *Harvard business review*, 79(3), 80-91.
55. Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio books.
56. Müller, M. Automated Well-Being Prediction Toolkit: Facial Analysis System, Visual Feature Extraction, PERMA Prediction. Available online: <https://github.com/mo12896/facial-analysis-system> (accessed on 26 May 2024).
57. Zeulner T, Hagerer GJ, Müller M, Vazquez I, Gloor PA. (2024) Predicting Individual Well-Being in Teamwork Contexts Based on Speech Features. *Information*. 15(4):217. <https://doi.org/10.3390/info15040217>
58. Coello, C. A. C. (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer.
59. Deb, K. (2011). Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing* (pp. 3-34). Springer. https://doi.org/10.1007/978-1-4471-2421-0_1
60. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241-251. <https://doi.org/10.1111/1469-7610.00715>
61. Pope, N. G. (2016). How the Time of Day Affects Productivity: Evidence from School Schedules. *The Review of Economics and Statistics*, 98(1), 1-11. https://doi.org/10.1162/REST_a_00525
62. Müller M, Dupuis A, Zeulner T, Vazquez I, Hagerer J, Gloor PA. (2024) Predicting Team Well-Being through Face Video Analysis with AI. *Applied Sciences*; 14(3):1284. <https://doi.org/10.3390/app14031284>
63. Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*. <https://doi.org/10.48550/arXiv.1811.12808>
64. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830. <https://doi.org/10.5555/1953048.2078195>
65. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673-690. <https://doi.org/10.1007/s11135-006-9018-6>
66. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 47(5), 1287-1294. <https://doi.org/10.2307/1911963>
67. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4), 817-838. <https://doi.org/10.2307/1912934>
68. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.
69. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1), 50-60. <https://doi.org/10.1214/aoms/1177730491>
70. Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
71. Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, 11-IT. <https://doi.org/10.2466/11.IT.3.1>
72. Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
73. Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publications.
74. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723. <https://doi.org/10.1109/TAC.1974.1100705>

75. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261-304. <https://doi.org/10.1177/0049124104268644>
76. Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
77. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar), 1157-1182. <https://doi.org/10.1162/153244303322753616>
78. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
79. Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (2020). Ngboost: Natural gradient boosting for probabilistic prediction. *Proceedings of the 36th International Conference on Machine Learning*, 2690-2700. <https://doi.org/10.48550/arXiv.1910.03225>
80. Liu, S., Wang, L., Lin, S., Yang, Z., & Wang, X. (2017). Analysis and prediction of team performance based on interaction networks. 2017 36th Chinese Control Conference (CCC), 4152-4156. <https://doi.org/10.23919/ChiCC.2017.8027933>
81. Murray, G., & Oertel, C. (2018). Predicting Group Performance in Task-Based Interaction. *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 14-21. <https://doi.org/10.1145/3242969.3243027>
82. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323. <https://doi.org/10.1145/331499.331504>
83. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R* (Vol. 112). Springer. <https://doi.org/10.1007/978-1-4614-7138-7>
84. Lundberg, S. M., Erion, G. G., & Lee, S. I. (2017). Consistent individualized feature attribution for tree ensembles. *Advances in Neural Information Processing Systems*, 30. <https://doi.org/10.48550/arXiv.1802.03888>
85. Marcílio, W., et al. (2020). Explainable artificial intelligence for understanding lung cancer survival. *Bioinformatics*, 36(9), 2904-2911. <https://doi.org/10.1093/bioinformatics/btaa060>
86. Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28), 307-317. <https://doi.org/10.1515/9781400881970-018>
87. Hair, J. F. (2009). *Multivariate data analysis*. Prentice Hall.
88. Küller, R., Ballal, S., Laike, T., Mikellides, B., & Tonello, G. (2006). The impact of light and colour on psychological mood: a cross-cultural study of indoor work environments. *Ergonomics*, 49(14), 1496-1507. <https://doi.org/10.1080/00140130600858142>
89. Tausczik, Y. R., & Pennebaker, J. W. (2013). Improving teamwork using real-time language feedback. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 459-468. <https://doi.org/10.1145/2470654.2470725>
90. Bellman, R. E. (1966). *Dynamic Programming*. Princeton University Press.
91. Page, P., Karaus, K., & Donner M. (2021) Enhancing Quality of Virtual Meetings through Facial and Vocal Emotion Recognition. Working Paper COIN Seminar, University of Cologne. https://www.dropbox.com/s/nftikcxaexls4y8/Team2_FinalPaper_main_Moody1.pdf?dl=0
92. Fiedler, J. Küpper, C., Schulz, J. & Weisshaar S. (2021) Building an Online-Dashboard to help presenters improve their remote meetings. Working Paper COIN Seminar, University of Cologne. https://www.dropbox.com/s/6px7ozmevtnezsm/Team3_Project%20Paper%20Moody%202.pdf?dl=0
93. G. Lordos & A. Lordos, A. (2019) Star city: Designing a settlement on mars. *Proceedings of the 22nd Annual Mars Society Convention*, Oct 2019, Los Angeles, CA
94. Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
95. Porter, B., & Grippa, F. (2020). A platform for AI-enabled real-time feedback to promote digital collaboration. *Sustainability*, 12(24), 10243.
96. Rietsche, R., Aier, S., & Michael, R. (2021). Does Real-Time Feedback Matter? A Simulation Study to Link Individual and Organizational Performance. WITS, the Workshop on Information Technologies and Systems.

97. Gloor, P. (2022). Happimetrics: Leveraging AI to Untangle the Surprising Link Between Ethics, Happiness and Business Success. Edward Elgar Publishing. <https://doi.org/10.4337/9781803924021>.
98. Gloor, P., Fischbach, K., Gluesing, J., Riopelle, K., & Schoder, D. (2018). Creating the collective mind through virtual mirroring based learning. *Development and Learning in Organizations: An International Journal*, 32(3), 4-7.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.