2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #52089

Exploring Team Dynamics and Cognitive Workload in Mixed Reality Collaborative Learning Environments

Khalid Bello, University of Louisville

Khalid Bello is a graduate student at University of Louisville.

Dr. Israa Azzam, Milwaukee School of Engineering

Dr. Azzam is an Assistant Professor of Mechanical Engineering at the Milwaukee School of Engineering (MSOE). Her teaching philosophy centers on creating an inclusive, learner-focused environment that emphasizes experiential learning, teamwork, and innovation. Dr. Azzam's goal is to integrate multiple teaching techniques and state-of-the-art methods to help students succeed. A major part of her approach is using advanced technologies such as Extended Reality (XR) in lectures and labs. She aims to bridge the gap between theory and practice, equipping students with essential STEM skills to thrive as innovators and future leaders. Dr. Azzam earned her Ph.D. in Mechanical Engineering Technology from Purdue University in 2025, where she specialized in digital technologies and control systems. Her doctoral research focused on industrial teleoperations, XR technology, and engineering education, and has been recognized with multiple awards. Before pursuing her Ph.D., Dr. Azzam completed her B.S. in Mechanical Engineering at Beirut Arab University (2019) and her M.E. in Mechanical Engineering at the American University of Beirut (2021). Her graduate work at AUB provided a strong theoretical foundation in control systems, which she expanded into applied research at Purdue, where she collaborated with industry partners on simulation-based and XR-integrated modules for engineering education.

Dr. Farid El Breidi, Purdue University - Purdue Polytechnic Institute – West Lafayette

Dr. Farid El Breidi is an Assistant Professor in the School of Engineering Technology at Purdue University. Dr. El Breidi is the director of the Digital Technologies Lab and is actively engaged in research encompassing digital fluid power, digital reality, and engineering pedagogy. He serves as the Principal Investigator for numerous internally and externally funded research projects. These projects have received support from agencies and associations such as the National Science Foundation and the National Fluid Power Association.

Dr. Faisal Aqlan, University of Louisville

Dr. Faisal Aqlan is an Associate Professor of Industrial Engineering at The University of Louisville. He received his Ph.D. in Industrial and Systems Engineering form The State University of New York at Binghamton.

Exploring Team Dynamics and Cognitive Workload in Mixed Reality Collaborative Learning Environments

Abstract

As mixed reality (MR) technologies become increasingly integrated into engineering education, understanding how teams experience and manage cognitive workload in immersive collaborative environments is critical. Although MR systems offer interactive and engaging learning experiences, most existing designs primarily emphasize individual users but neglect the complex coordination processes and shared cognitive demands that emerge during team-based tasks. Further, limited empirical work has examined how physiological indicators of workload interact with team processes and usability perceptions at the team level in MR contexts. This constrains the development of adaptive systems that effectively support collaborative performance. To address these gaps, this study investigates team dynamics and cognitive workload during a collaborative MR learning task. 33 teams of engineering students completed a hydraulic bike assembly task using an MR platform while multimodal data were collected, including heart rate (HR), heart rate variability (HRV), perceived system usability, and teamwork assessments. Unsupervised cluster analysis was applied to physiological, teamwork, and usability indicators to identify distinct team experience profiles, which were subsequently examined in relation to self-reported cognitive workload. The analysis revealed two distinct team profiles. One cluster was characterized by higher physiological synchrony, stronger teamwork behaviors, and more positive usability perceptions, whereas the other cluster exhibited lower physiological synchrony, weaker teamwork and usability evaluations. Although differences in self-reported cognitive workload between the two clusters did not consistently reach conventional levels of statistical significance, effect size estimates indicated a moderate tendency toward higher workload in one cluster. These findings suggest that variations in physiological organization, teamwork quality, and usability perceptions are meaningfully associated with how teams experience cognitive demands in MR environments. This study advances understanding of team-level adaptation in immersive collaborative learning contexts and highlights the value of integrating physiological and perceptual measures to characterize team experiences.

Keywords: Mixed Reality; Team Dynamics; Cognitive Workload; Heart Rate; Heart Rate Variability; Collaborative Learning; Cluster Analysis

1. Introduction

Engineering education increasingly emphasizes experiential and collaborative learning to better prepare students for professional practice. Many engineering tasks such as system assembly, troubleshooting, and design require teams to coordinate actions, share information, and jointly manage complex problem spaces. However, supporting effective collaboration in educational settings remains challenging, particularly when tasks involve high cognitive demands and dynamic interactions among team members. Understanding how teams experience and manage these demands is therefore a critical concern for the design of learning environments that aim to foster both technical competence and collaborative skills.

A family of technologies that has been increasingly explored to support such hands-on, collaborative learning is immersive technologies, which can be situated along the reality-virtuality continuum proposed by Milgram and Kishino [1]. This continuum spans a range of experiences from fully physical to fully virtual environments, and provides a useful framework for distinguishing virtual reality (VR), augmented reality (AR), and mixed reality (MR). VR sits at the virtual end of the continuum, fully immersing users in computer-generated environments that replace the physical world and are typically experienced through head-mounted displays [1]. AR, in contrast, overlays digital content such as text, images, or three-dimensional models onto the user's view of the real world, and is commonly characterized by the combination of real and virtual content, real-time interaction, and registration of virtual elements in three dimensions [2] MR extends this idea further by anchoring virtual content to the physical environment in a spatially aware manner, so that digital and physical elements can be perceived and manipulated as if they coexisted in the same space [3]. Among these technologies, MR has gained. particular attention as a promising approach for addressing the challenges of collaborative engineering education [4]. By integrating virtual content with physical environments, MR allows learners to interact with digital representations while performing real-world tasks. In engineering education, MR has been used to support activities such as mechanical assembly, procedural training, and spatial visualization [5]. MR environments can reduce abstraction, provide contextualized guidance, and enable hands-on practice [6]. While such affordances have been shown to benefit individual learners, their implications for collaborative learning are less well understood [7].

One key factor influencing the effectiveness of collaborative learning in MR environments is cognitive workload [8]. Cognitive workload reflects the mental effort required to perform tasks, shaped by their complexity and the demands placed on the brain's limited processing resources [9]. In team-based MR activities, learners must not only attend to the task and the virtual interface, but also communicate with teammates, monitor shared progress, and adapt to others' actions in real time. These additional coordination demands can significantly alter how workload is experienced and distributed across a team. When workload is poorly managed, collaboration may become inefficient or error-prone; when appropriately balanced, it can support effective teamwork and learning [10]. Despite its importance, cognitive workload in MR-based learning has largely been examined from an individual perspective. Existing studies often rely on self-reported workload measures or focus on isolated performance metrics, offering limited insight into how teams collectively experience and manage cognitive demands. For instance, Michalski, et al. [11] examined cognitive load in a MR heart anatomy application for first year medical students, relying solely on the Cognitive Load Component questionnaire to capture intrinsic, extraneous, and germane load at the individual level. Similarly, Mugisha and Arguel [12] investigated cognitive load during arthroscopic knot tying using HoloLens 2, administering self-report scales to each learner. Even when MR systems are deployed in collaborative or instructional settings, workload is typically aggregated by averaging individual self-reports [13], which may obscure substantial variability in how different teams interact with the system, perceive usability, and coordinate their actions. This lack of attention to team-level variability limits our understanding of why some teams perform effectively in MR environments while others struggle.

Recent developments in physiological sensing provide new opportunities to examine cognitive workload in collaborative settings. Measures such as heart rate (HR) and electrodermal activity (EDA) offer continuous, objective indicators of cognitive and emotional activation during task performance [14]. When combined with perceptual measures such as system usability and teamwork assessments, these data can provide a richer picture of how teams experience MR environments. However, to our knowledge, research integrating physiological and perceptual indicators to study team dynamics and cognitive workload in collaborative MR learning remains scarce. Another challenge lies in how such multimodal data are analyzed. Traditional analytical approaches often assume homogeneous user experiences and linear relationships between variables. For example, variable centered methods such as ANOVA, multiple regression, and structural equation modeling are commonly used in MR and collaborative learning research, but they emphasize generalizable insights across a population while overlooking individual heterogeneity [15]. Törmänen, et al. [16] explicitly note that prior research on collaborative learning has predominantly relied on variable centered approaches, which examine averaged relations between variables and assume that all learners or teams experience the process in the same way. In collaborative learning contexts, however, teams may exhibit distinct patterns of physiological activation, usability perception, and coordination behavior. Data-driven techniques, such as cluster analysis, allow for the identification of meaningful groupings based on similarities across multiple indicators, enabling the characterization of different team experience profiles rather than relying solely on aggregate trends [17, 18].

This study examines how teams engage in a collaborative MR learning task. By collecting physiological data alongside perceived system usability and teamwork assessments, and by applying cluster analysis at the team level, the study seeks to identify distinct patterns of team experience. Through this approach, the study aims to contribute to a deeper understanding of team dynamics in immersive collaborative learning environments and to inform the design of MR systems that better support effective collaboration and balanced cognitive demands.

The remainder of this paper is organized as follows. Section 2 discusses relevant research. Section 3 describes the collaborative MR learning environment, experiment, data collection and analysis methods, including physiological measures and clustering techniques. Section 4 presents the results, and Section 5 discusses the implications and conclusions of the study.

2. Relevant Research

MR has been widely studied as a learning and training technology for tasks that require spatial reasoning, procedural execution, and interaction with physical artifacts. In manufacturing assembly and maintenance contexts, MR systems commonly present spatially registered overlays, step wise instructions, and interactive cues to support accurate execution and reduce reliance on external documentation. Azzam, et al. [19] established the feasibility of MR for hands-on technical learning and highlighted that learning and performance benefits are strongly shaped by interface design. More recently, the focus of MR-based learning has extended from single-user training to collaborative MR environments in which multiple learners share a workspace and coordinate on a common task. For example, Choi, et al. [4] developed a collaborative MR learning platform for mechanical engineering concepts and emphasized the importance of shared interaction when teams complete assembly oriented tasks. Wang and Dunston [20] introduced MR-based collaborative virtual environments for design review and demonstrated their effectiveness in enhancing spatial comprehension, reducing performance time, and improving mental interpretation during design error detection tasks. Although these studies demonstrate the promise of collaborative MR for engineering learning, collaboration introduces additional demands related to coordination and shared task management that warrant targeted investigation.

Cognitive workload is a central construct in learning and human factors research because it captures the mental effort required to perform a task under specific conditions. Cognitive Load Theory (CLT) explains how limited working memory resources constrain learning and performance, and it distinguishes between load imposed by the task itself, and load introduced by instructional or interface design [21]. In immersive environments, workload may be shaped by sensory richness, interaction mechanics, and the need to integrate virtual and physical information sources. Evidence from virtual reality (VR) research suggests that collaborative tasks in immersive settings can increase extraneous load, particularly when interface elements and environmental features compete for attention [22]. Kirschner, et al. [10] argued that collaboration can be beneficial for complex tasks because team members may distribute cognitive demands across a collective working memory. At the same time, collaboration introduces coordination and communication costs that can offset these benefits when group processes are inefficient. Subsequent work has applied collaborative cognitive load theory to computer supported collaborative learning and has shown that the effectiveness of collaboration depends on the balance between interactivity benefits and coordination costs [23]. These ideas are especially relevant to collaborative MR learning because immersive interfaces can both support shared reference and increase coordination demands, making team workload an emergent property of task design, system usability, and teamwork processes.

Most studies assess cognitive workload using validated self-report instruments, both because they are practical and because they capture the learner's experienced demand. The NASA Task Load Index (NASA-TLX) is one of the most widely used tools in this area and operationalizes workload across multiple dimensions, including mental demand, temporal demand, effort, and frustration [24]. In simulated environments, such as extended reality (XR) settings, the Simulation Task Load Index (SIM-TLX) has been developed to measure workload within these contexts [25]. The SIM-TLX has been used across a wide range of task settings and remains useful for summarizing perceived workload after task completion, although it does not provide moment to moment information during dynamic activity. System usability is commonly assessed using the System Usability Scale (SUS), a ten-item instrument designed to provide a global usability assessment across diverse technologies. Brooke [26] introduced SUS as a quick and reliable usability measure, and subsequent work has supported its broad applicability in system evaluation. In collaborative MR learning, usability is not only an experience measure but also a potential driver of workload because interaction friction can consume cognitive resources that would otherwise be available for task reasoning and team coordination [27]. This link motivates including usability measures when studying workload in immersive collaborative environments.

Since subjective measures are typically collected after a task and can miss fluctuations during performance, studies incorporate physiological data to infer cognitive and affective states continuously [28]. Recent work continues to demonstrate that EDA and HRV features can differentiate task conditions associated with cognitive load modulation and can align with self-report workload patterns [29]. Reviews of non-intrusive workload assessment methods also emphasize the growing role of physiological signals for real time workload inference [30]. In educational contexts, multimodal studies have increasingly combined physiological measures with behavioral signals to model cognitive demands. For example, recent studies have integrated eye tracking, EDA, and heart related measures to examine relationships between attention, comprehension, and cognitive workload during learning tasks [31]. These results support the feasibility of using physiological indicators as cognitive workload relevant signals.

An important shift in collaborative learning analytics is the move from individual physiology to team level physiological dynamics. Interpersonal physiological synchrony has shown that physiological coupling can emerge during collaborative activities and that synchrony patterns can differ between independent work and group discussion [32]. Liu, et al. [33] collected EDA and HR in naturalistic classroom sessions and examined physiological synchrony as a potential marker of collaborative learning quality. Dindar, et al. [34] proposed methods for detecting shared physiological arousal events during collaborative problem-solving using EDA, providing a way to locate moments of collective challenge or alignment in team interaction. These studies suggest that physiological data can capture aspects of team interaction that are not easily observable through self-report alone. At the same time, team physiology is not a direct measure of teamwork quality or workload. Physiological similarity may reflect shared attention, shared stress, or common task phases. This ambiguity strengthens the case for integrating physiological measures with perceptual indicators of usability and teamwork processes when characterizing team experience in collaborative MR learning.

Multimodal learning analytics has grown rapidly as researchers seek to capture learning processes through combinations of physiological, behavioral, and contextual data [35]. Surveys and mapping studies have emphasized that multimodal approaches can provide richer accounts of learning and collaboration than single modality methods, while also highlighting challenges related to fusion methods, validity, and interpretability [36]. Within this broader area, clustering and other unsupervised methods have been used to identify distinct patterns of learner behavior and collaboration [37]. Xu, et al. [38] proposed a multimodal analytics framework to identify collaborative clusters in face-to-face learning settings and used cluster characterization to reveal differences in collaborative patterns. In a related direction, Acosta, et al. [39] examined how multimodal signals can support classification of collaboration quality or satisfaction, demonstrating that latent profiles can be detected from interaction traces and sensor data. However, applications of data driven profiling to collaborative MR learning remain limited, particularly when the goal is to characterize team experience profiles that jointly reflect physiological activation, usability perceptions, and teamwork processes. Existing MR studies have demonstrated collaborative platforms and have evaluated aspects of team behavior, but they often do not integrate physiology and perceptual measures into a unified profiling approach. This gap motivates a team centered, multimodal clustering approach that can reveal heterogeneity in how teams experience collaborative MR tasks.

3. Method and Analysis

3.1 Participants

A total of 103 students were recruited from an undergraduate course, with 100 of them completing the demographic questionnaire (90 identifying as male and 10 as female). Participant ages ranged primarily between 18 and 23 years, with 44 individuals aged 18–20, 51 aged 21–23, and five participants older than 23. Most participants reported little to no prior experience with mixed reality (MR). Before beginning the experimental tasks, all participants completed an introductory training session in the MR environment to become familiar with the system controls and object manipulation methods. Although all participants were drawn from the same course, team members were intentionally grouped with peers they did not usually collaborate with to reduce the effects of prior familiarity. After excluding data due to sensor connectivity issues and participants who did not consent to physiological data collection, the final sample consisted of 33 teams. Owing to the skewed gender distribution, most teams were composed entirely of male participants, with the few female participants allocated across teams when possible. The study protocol was approved by the Institutional Review Board (approval Number: IRB-2024–1403) at Purdue University.

3.2 Collaborative Mixed Reality Learning Environment

The MR activity requires teams to collaboratively assemble a hydraulic bike composed of four interrelated subsystems (see Figure 1). The task takes a total time of 13 minutes and is divided into three consecutive stages. Upon completion of each stage, an audio cue indicates the transition, the associated virtual components are removed from view, and participants must select a “Proceed” button to continue. During the initial five-minute stage, participants focus on assembling the tank-accumulator and gearbox subsystems. This stage presents the highest level of complexity, as it involves intricate hydraulic and mechanical elements that require close coordination among team members. The second stage, lasting four minutes, centers on the construction of the clutch and frame subsystems. While this phase is less complex than the first, it still demands sustained collaboration and attention. The final stage also spans four minutes and involves combining all previously assembled subsystems into a complete hydraulic bike. Additional elements, including wheels and cranks, are incorporated at this point, and the assembly must be finalized before the time limit expires.

Figure 1: Sample snapshots from the MR environment showing a participant assembling bike components

3.3 Data Collection and Preprocessing

Physiological signals were recorded continuously from all participants using EmotiBit wearable devices, which provide validated measures of HR and inter-beat intervals (IBI) for HRV analysis. Following task completion, participants completed post-task questionnaires assessing teamwork, system usability, and cognitive workload. HR and HRV signals were preprocessed to remove noise and artifacts prior to analysis. Outliers were identified using interquartile range criteria and corrected via linear interpolation. HR signals were denoised using wavelet-based filtering, and HRV was derived from cleaned inter-beat interval data. HRV time series were resampled at 1 Hz and normalised using z-score transformation to reduce inter-individual variability. Physiological synchrony within teams was quantified using dynamic time warping (DTW) applied separately to HR and HRV time series. DTW captures similarity between time series that may differ in timing or pacing, allowing assessment of shared physiological dynamics across team members. Lower DTW values indicated higher synchrony. To characterize the structure and complexity of team-level physiological dynamics, multidimensional recurrence quantification analysis (MdRQA) was applied to HR and HRV signals. MdRQA treats the team as a single multidimensional dynamical system and enables analysis of collective coordination beyond pairwise interactions. From recurrence plots, four recurrence metrics were extracted for both HR and HRV: determinism (DET), average diagonal line length (ADL), laminarity (LAM), and entropy (ENTR). These measures capture the predictability, temporal stability, and complexity of physiological dynamics within teams.

Teamwork processes were assessed using selected subscales of an adapted Team Assessment and Diagnostic Instrument (TADI) [40], focusing on task and communication skills (TCS), team resources and working environment (TRWE), and attitude toward teammates and task. All items were rated on Likert-type scales and aggregated to the team level. Perceived system usability was measured using the SUS. Team cognitive workload, the outcome variable of this study, was assessed using the SIM-TLX and aggregated at the team level. All physiological, teamwork, and usability variables were standardized using z-score transformation prior to clustering to ensure comparability across measures. Team cognitive workload was excluded from the clustering procedure and retained as an external outcome variable for validation and comparison across identified team profiles.

3.4 Clustering Method and Evaluation Metrics

To identify distinct team experience profiles in the mixed reality (MR) collaborative learning environment, K-means clustering was employed as the primary unsupervised learning method. K-means is a partition-based clustering technique that groups observations into a predefined number of clusters by minimizing within-cluster variance while maximizing separation between cluster centroids [41]. This method is widely used for exploratory analysis of multivariate data and is particularly suitable for identifying latent structure in datasets where no prior class labels are available. Before clustering, all variables included in the analysis were standardized using z-score transformation to ensure that features measured on different scales contributed equally to the Euclidean distance calculations used by the K-means algorithm. Clustering was conducted for a range of cluster numbers to examine alternative solutions and assess their quality and interpretability. As an initial step, the elbow method was applied to support the selection of an appropriate number of clusters [42]. This method examines the relationship between the number of clusters (k) and the within-cluster sum of squared distances (WCSS), which reflects cluster compactness. K-means clustering was performed across a range of k values, and the resulting WCSS values were plotted to identify a point at which the marginal reduction in WCSS diminished substantially. The final selection of the clustering solution was informed by both the elbow analysis and additional internal evaluation metrics. To comprehensively assess the quality of the clustering solutions, three internal validation indices were used. The Silhouette Score (SC) measures the degree to which each observation is similar to its assigned cluster relative to other clusters [43]. The score ranges from −1 to 1, with higher values indicating better defined clusters. A higher average silhouette score suggests that observations are well matched to their own cluster and poorly matched to neighboring clusters, reflecting a more coherent clustering structure. The Davies–Bouldin Index (DBI) evaluates clustering quality by considering the ratio of within-cluster dispersion to between-cluster separation. Lower DBI values indicate better clustering solutions, as they reflect compact clusters that are well separated from one another [44]. The Calinski–Harabasz Index (CHI), also known as the variance ratio criterion, assesses clustering quality based on the ratio of between-cluster variance to within-cluster variance. Higher CHI values indicate superior clustering performance, reflecting greater separation between clusters relative to internal dispersion [45].

4. Results

4.1 Clustering and Team Profiles

To determine an appropriate number of clusters, solutions ranging from two to six clusters were examined using the elbow method and multiple internal validation indices, including the Silhouette Score (SC), Davies–Bouldin Index (DBI), and Calinski–Harabasz Index (CHI). Inspection of the elbow plot revealed a marked reduction in WCSS from two to three clusters, followed by diminishing returns as additional clusters were added **(see Figure 2)**. This pattern suggested that a small number of clusters captured most of the variance in the data. In this plot, the x-axis represents the number of clusters (k) and the y-axis represents the within-cluster sum of squared distances (WCSS), which measures how compactly grouped the data points are within each cluster. As k increases, WCSS decreases because more clusters allow tighter groupings. The "elbow" is the point where the rate of decrease sharply levels off, indicating that adding more clusters yields diminishing improvements in fit. The optimal k is identified at this inflection point, balancing model complexity against explanatory power.

Consistent with this observation, the two-cluster solution demonstrated the highest Silhouette Score (SC = 0.22) and the highest Calinski–Harabasz Index (CHI = 10.46) among the evaluated solutions **(see Table 1)**, indicating superior cluster separation and compactness. Although the Davies–Bouldin Index decreased slightly with higher cluster numbers, the two-cluster solution provided the most favourable overall balance across the three metrics. Based on these results and the interpretability of the resulting profiles, a two-cluster solution was selected for further analysis.

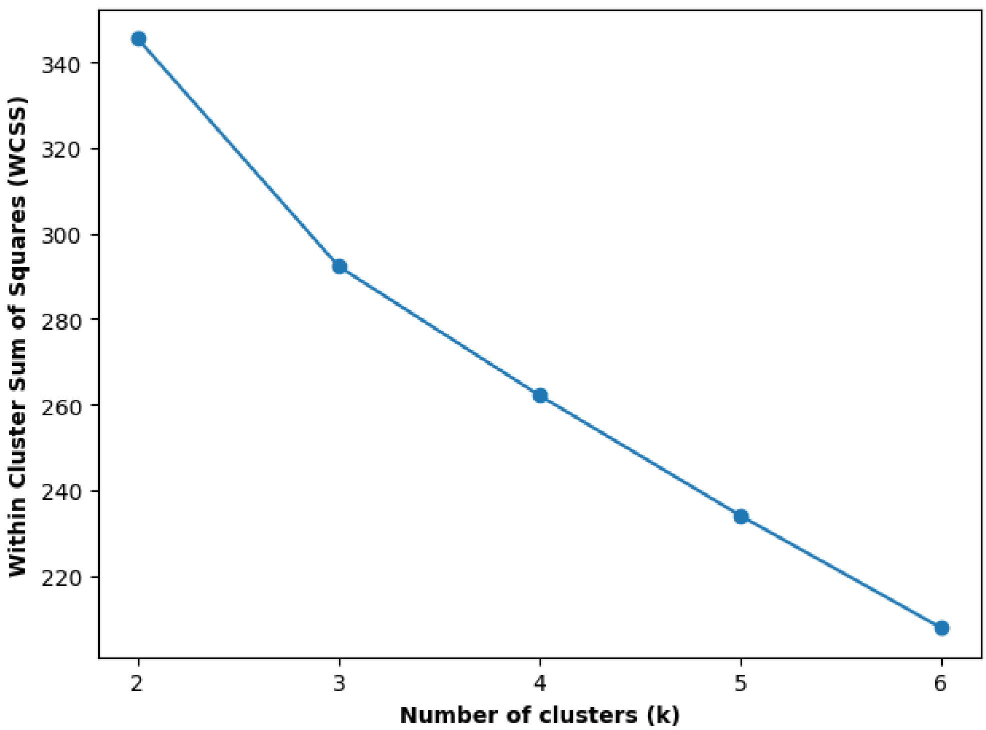

Figure 2: Optimal number of clusters using the Elbow method

Table 1: K-means internal validation metrics across k

k-clusters/ Clustering metrics	**2**	**3**	**4**	**5**	**6**
Silhouette index	0.22	0.15	0.14	0.14	0.16
Davies-Boudin index	1.59	1.53	1.65	1.43	1.30
Calinski-Harabasz index	10.46	8.71	7.37	6.81	6.60

To visualize the distribution of the identified clusters, principal component analysis (PCA) was applied to reduce the dimensionality of the standardized feature space prior to plotting. **Figure 3** shows the projection of teams onto the first two principal components, illustrating the arrangement of physiological, teamwork, and usability features under the two-cluster solution. The centroids of the two clusters were examined using cluster mean profiles of z-scored features, which are presented in **Figure 4**. The profiles revealed systematic differences between clusters across physiological recurrence metrics, teamwork indicators, and perceived system usability.

Cluster 1 consisted of 22 teams and was characterized by higher values across most HR and HRV recurrence-based measures. Higher DET and LAM in this cluster indicate more predictable, stable, and structured physiological dynamics, reflecting a greater tendency for the system to follow deterministic trajectories and remain in similar states over time. In addition, higher average diagonal line length (ADL) suggests sustained recurrence patterns, indicating longer-lasting temporal coherence in physiological activity. Collectively, these features describe a physiologically organized yet dynamically rich team state, characterized by both stability and adaptive complexity.

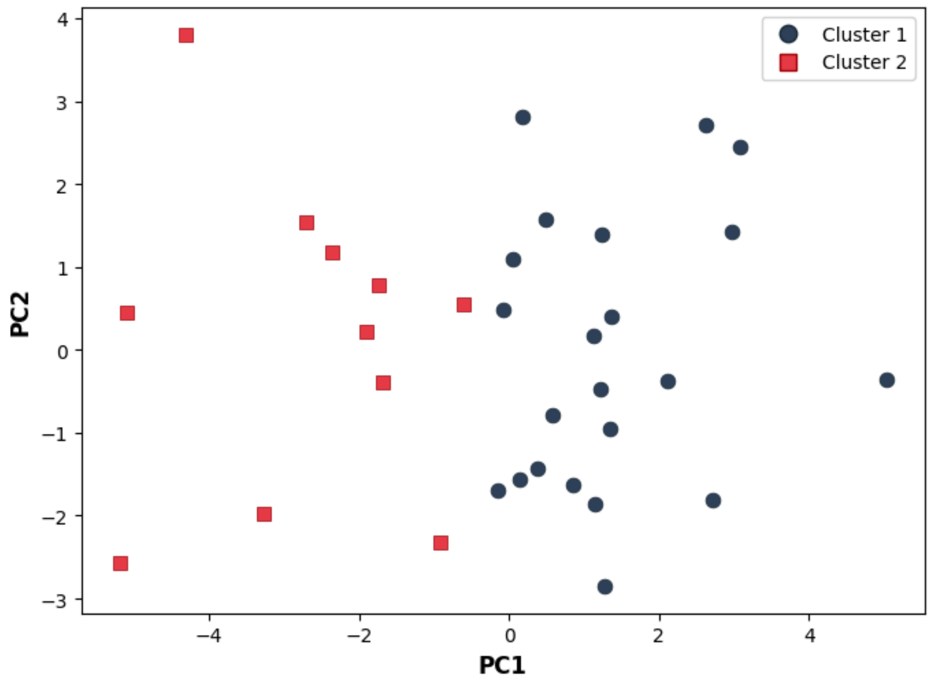

Figure 3: Data distribution into 2 clusters

In addition to physiological differences, teams in Cluster 1 reported substantially higher scores on teamwork-related measures, including TCS, TRWE, and attitudes toward teammates and the task. System usability scores were also higher for this cluster, indicating more positive perceptions of the MR platform. Phase synchronization measures based on DTW tended to be lower in this cluster. Given that lower DTW values correspond to higher synchrony, this pattern suggests less differentiated and more tightly synchronized physiological responses among team members.

Cluster 2, comprising 11 teams, exhibited a contrasting profile. Teams in this cluster showed lower values on most HR and HRV recurrence metrics, including DET, LAM, ADL, and ENTR, indicating less predictable, less stable, and less complex physiological dynamics. Shorter diagonal structures and reduced laminarity suggest weaker temporal coherence and fewer prolonged stable states in physiological activity. At the same time, Cluster 2 displayed higher DTW values, reflecting lower physiological synchrony within teams. From a behavioral perspective, Cluster 2 was characterized by lower teamwork and usability scores, including reduced TCS, less favorable working environments, and more negative attitudes toward teammates and the task. System usability ratings were also lower, suggesting that teams in this cluster experienced greater difficulty interacting with the MR system. These results indicate that the two clusters represent distinct team experience profiles in the MR collaborative learning context. Cluster 1 reflects teams with well-organized, temporally coherent, and dynamically rich physiological patterns, combined with strong teamwork behaviors and positive usability perceptions. In contrast, Cluster 2 represents teams with less structured and less complex physiological dynamics, lower teamwork quality, poorer usability perceptions, and reduced physiological synchrony.

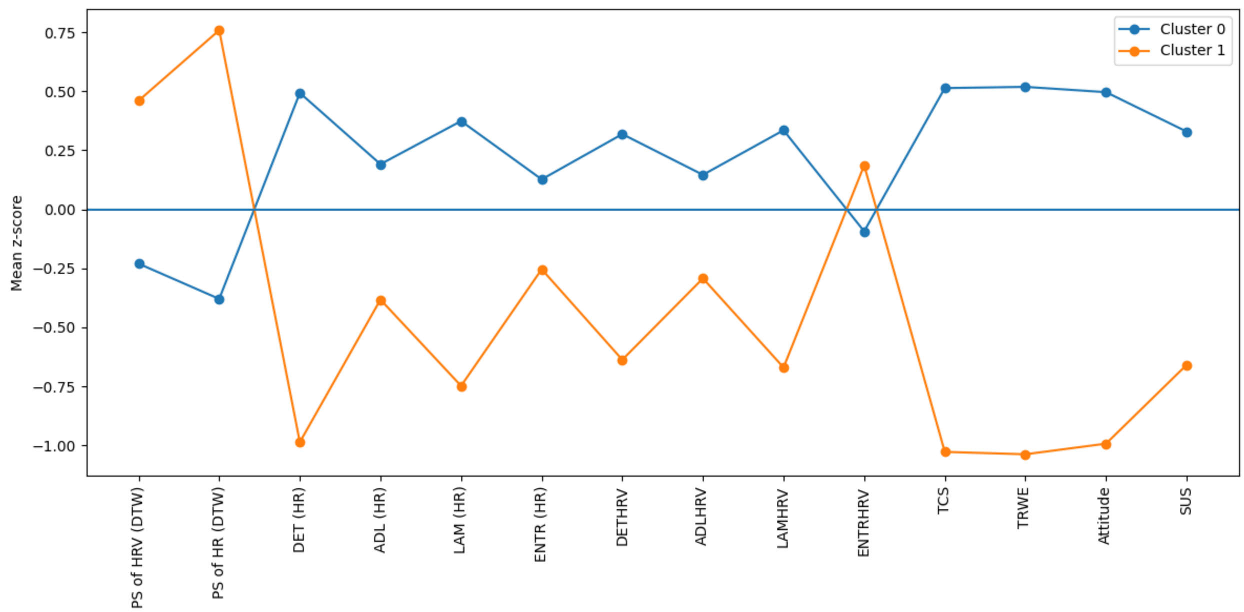

Figure 4: Cluster mean profile (Z-score)

4.2 Difference in Team Cognitive Workload between Clusters

Descriptive statistics indicated that Cluster 2 ($n = 11$) reported a higher mean workload ($M = 2.27$, $SD = 0.37$) than Cluster 1 ($n = 22$), which exhibited a lower average workload ($M = 2.05$, $SD = 0.23$). Prior to inferential testing, the assumptions of normality and homogeneity of variance were assessed. Shapiro–Wilk tests did not indicate departures from normality within either cluster ($ps > .41$). Levene's test for equality of variances was not statistically significant ($p = .065$), although the result was borderline, suggesting potential variance heterogeneity given the unequal cluster sizes. Because the clustering solution comprised two groups, workload differences were evaluated using independent-samples comparisons. While a t-test assuming equal variances indicated a statistically significant difference between clusters ($p = .043$), the more robust Welch t-test ($p = .090$) and the nonparametric Mann–Whitney U test ($p = .126$) did not reach statistical significance. Accordingly, the difference in workload between clusters should be interpreted as a non-significant trend, with Cluster 2 exhibiting moderately higher workload than Cluster 1.

Effect size analyses provided additional insight into the magnitude of the observed difference. The comparison yielded a moderate-to-large negative Hedges' g ($g = −0.76$), indicating higher workload in Cluster 2 relative to Cluster 1, with a 95% confidence interval that narrowly excluded zero [−1.70, −0.01]. In contrast, Cliff's delta indicated a small-to-moderate effect ($\delta = −0.34$), with a confidence interval that included zero [−0.73, 0.11], suggesting uncertainty in the robustness of the effect. These results indicate a trend toward higher cognitive workload in Cluster 2, with effect size estimates suggesting a potentially meaningful difference; however, the lack of consistent statistical significance across robust tests warrants cautious interpretation.

5. Discussion and Conclusions

This study investigated how teams experience cognitive workload in a collaborative MR learning environment by integrating physiological dynamics, teamwork processes, and usability perceptions at the team level. Using K-means clustering, two distinct team experience profiles were identified, revealing systematic differences in physiological organization, coordination quality, and subjective perceptions of teamwork and system usability. Cluster 1, comprising the majority of teams, was characterized by higher values across recurrence-based physiological metrics, including determinism, laminarity, and average diagonal line length. These features indicate more predictable, stable, and temporally coherent physiological dynamics, suggesting that teams in this cluster maintained structured and sustained coordination patterns during the collaborative task. This physiological profile was accompanied by stronger teamwork processes and more positive usability perceptions, suggesting that well-organized physiological dynamics co-occur with effective collaboration and smoother interaction with the MR system. In contrast, the second cluster exhibited lower recurrence structure and complexity, alongside higher DTW-based synchronization distances, indicating weaker temporal coherence and reduced physiological coordination within teams. Behaviorally, these teams reported poorer teamwork quality and lower system usability. Together, these findings suggest that less structured physiological organization may reflect challenges in coordination, communication, or task management in immersive collaborative environments.

Differences in team cognitive workload across clusters further illuminate the relationship between these team profiles and subjective experience. Although teams in Cluster 2 reported higher mean workload than those in Cluster 1, inferential tests yielded mixed results. While a standard t-test suggested a statistically significant difference, more robust analyses accounting for unequal variances (Welch's t-test) and distributional assumptions (Mann–Whitney U test) did not reach conventional levels of significance. Effect size analyses provided additional context for interpreting this pattern. The moderate-to-large Hedges' g suggests that the workload difference may be practically meaningful, whereas Cliff's delta, with a confidence interval spanning zero, indicates uncertainty in the robustness of this effect. These findings suggest that teams characterized by less organized physiological dynamics, weaker teamwork processes, and poorer usability perceptions may experience higher cognitive demands, even if these differences are not consistently detectable through inferential tests in the present sample. These results have important implications for the design and evaluation of collaborative MR learning environments. They suggest that supporting teams in maintaining organized, coherent physiological dynamics, potentially through improved interface design, clearer task structuring, or adaptive feedback mechanisms may help mitigate excessive cognitive workload and enhance collaborative effectiveness. Moreover, the integration of physiological measures with subjective assessments provides a richer understanding of team experiences that cannot be captured through self-report alone.

Several limitations should be acknowledged. First, the relatively small number of teams limits statistical power, particularly for detecting workload differences between clusters. Second, the skewed gender distribution restricts the generalizability of the findings to more diverse populations. Future research should examine larger and more diverse sample and explore causal relationships between physiological dynamics, teamwork processes, and performance outcomes. Longitudinal designs and experimental manipulations of MR interface features may further clarify how adaptive systems can dynamically support team coordination and workload regulation. This study demonstrates that teams engaging in collaborative MR learning tasks can be meaningfully differentiated based on integrated physiological, behavioral, and perceptual indicators. Teams with more organized and temporally coherent physiological dynamics exhibited stronger teamwork, more positive usability perceptions, and a tendency toward lower cognitive workload. Conversely, teams with less structured physiological organization and poorer collaboration experienced greater cognitive demands. Although workload differences did not consistently reach statistical significance, effect size estimates suggest potentially meaningful trends. These findings advance understanding of team-level adaptation in immersive learning environments and highlight the value of combining physiological and subjective measures to inform the design of adaptive MR systems that better support collaborative learning and cognitive balance.

References

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems,* vol. 77, no. 12, pp. 1321-1329, 1994.

[2] R. T. Azuma, "A survey of augmented reality," *Presence: teleoperators & virtual environments,* vol. 6, no. 4, pp. 355-385, 1997.

[3] M. Speicher, B. D. Hall, and M. Nebeling, "What is mixed reality?," in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1-15.

[4] M. Choi, Y. Jiang, F. Breidi, C. Mousas, and M. Akdere, "A mixed reality platform for collaborative technical assembly training," in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, 2022, pp. 1-2.

[5] I. Azzam, F. Breidi, F. Aqlan, J. Garcia, and P. Asunda, "Mixed Reality as a Teaching Tool for Improving Spatial Visualization in Engineering Students," 2024: ASEE Conferences.

[6] A. Bondin *et al.*, "Rethinking Engineering Education in Light of Mixed Reality Opportunities–Results from a European Survey," *Procedia computer science,* vol. 277, pp. 2554-2563, 2026.

[7] A. A. Ali, G. A. Dafoulas, and J. C. Augusto, "Collaborative educational environments incorporating mixed reality technologies: A systematic mapping study," *IEEE Transactions on Learning Technologies,* vol. 12, no. 3, pp. 321-332, 2019.

[8] Y. Hou, Q. Xie, N. Zhang, and J. Lv, "Cognitive load classification of mixed reality human computer interaction tasks based on multimodal sensor signals," *Scientific Reports,* vol. 15, no. 1, p. 13732, 2025.

[9] K. Stojmenova and J. Sodnik, "Methods for assessment of cognitive workload in driving tasks," in *ICIST 2015 5th International Conference on Information Society and Technology*, 2015, pp. 229-234.

[10] P. A. Kirschner, J. Sweller, F. Kirschner, and J. Zambrano R, "From cognitive load theory to collaborative cognitive load theory," *International journal of computer-supported collaborative learning,* vol. 13, no. 2, pp. 213-233, 2018.

[11] C. Michalski, M. Cowan, J. Bohinsky, R. Dickerson, and J. H. Plochocki, "Evaluation of cognitive load for a mixed reality anatomy application," *Translational Research in Anatomy,* vol. 31, p. 100247, 2023.

[12] G. Mugisha and A. Arguel, "Procedural learning in mixed reality: Assessing cognitive load and performance," *Education Sciences,* vol. 15, no. 3, p. 339, 2025.

[13] W. S. Helton, G. J. Funke, and B. A. Knott, "Measuring workload in collaborative contexts: Trait versus state perspectives," *Human factors,* vol. 56, no. 2, pp. 322-332, 2014.

[14] H. F. Posada-Quintero and J. B. Bolkhovsky, "Machine learning models for the identification of cognitive tasks using autonomic reactions from heart rate variability and electrodermal activity," *Behavioral Sciences,* vol. 9, no. 4, p. 45, 2019.

[15] S. López-Pernas, H. Kayaduman, L. V. Vogelsmeier, and M. Saqr, "The three levels of analysis: Variable-centered, person-centered and person-specific analysis in education," in *Advanced learning analytics methods: AI, precision and complexity*: Springer, 2025, pp. 493-519.

[16] T. Törmänen, H. Järvenoja, M. Saqr, J. Malmberg, and S. Järvelä, "A person-centered approach to study students' socio-emotional interaction profiles and regulation of collaborative learning," in *Frontiers in Education*, 2022, vol. 7: Frontiers Media SA, p. 866612.

[17] J. W. You, "Relationship between team learning profiles and outcomes in team project-based learning: a cluster analysis," *Studies in Higher Education,* vol. 49, no. 1, pp. 16-32, 2024.

[18] L. A. d. Souza and H. G. Costa, "Clustering members of project teams according their perceptions about the impact of remote work on project success: a comparative analysis," *Pesquisa Operacional,* vol. 43, p. e276911, 2023.

[19] I. Azzam, K. Bello, F. Breidi, and F. Aqlan, "Analysis of User Experience in Digital Reality: A Comparative Study of VR and MR for Manufacturing Training," in *2025 ASEE Annual Conference & Exposition*, 2025.

[20] X. Wang and P. S. Dunston, "Comparative effectiveness of mixed reality-based virtual environments in collaborative design," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 41, no. 3, pp. 284-296, 2011.

[21] F. Paas and J. J. Van Merrienboer, "Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks," *Current Directions in Psychological Science,* vol. 29, no. 4, pp. 394-398, 2020.

[22] A. Queiroz *et al.*, "Collaborative tasks in immersive virtual reality increase learning," in *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning-CSCL 2023, pp. 27-34*, 2023: International Society of the Learning Sciences.

[23] C. Kent and M. Cukurova, "Investigating collaboration as a process with theory-driven learning analytics," *Journal of Learning Analytics,* vol. 7, no. 1, pp. 59-71, 2020.

[24] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, 2006, vol. 50, no. 9: Sage publications Sage CA: Los Angeles, CA, pp. 904-908.

[25] D. Harris, M. Wilson, and S. Vine, "Development and validation of a simulation workload measure: the simulation task load index (SIM-TLX)," *Virtual Reality,* vol. 24, no. 4, pp. 557-566, 2020.

[26] J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry,* vol. 189, no. 194, pp. 4-7, 1996.

[27] H. M. Cuevas, S. M. Fiore, B. S. Caldwell, and L. Strater, "Augmenting team cognition in human-automation teams performing in complex operational environments," *Aviation, space, and environmental medicine,* vol. 78, no. 5, pp. B63-B70, 2007.

[28] A. Darzi, T. Wondra, S. McCrea, and D. Novak, "Classification of different cognitive and affective states in computer game players using physiology, performance and intrinsic

factors," in *International Conference on Intelligent Human Systems Integration*, 2019: Springer, pp. 23-29.
[29] A. Boffet, L. M. Arsac, V. Ibanez, F. Sauvet, and V. Deschodt-Arsac, "Detection of Cognitive Load Modulation by EDA and HRV," *Sensors,* vol. 25, no. 8, p. 2343, 2025.
[30] N. Basnet and M. Zahabi, "Real-time cognitive workload assessment using non-intrusive methods: a systematic review," *Ergonomics,* pp. 1-26, 2025.
[31] Y. Que, Y. Zheng, J. H. Hsiao, and X. Hu, "Using eye movements, electrodermal activities, and heart rates to predict different types of cognitive load during reading with background music," *Scientific Reports,* vol. 15, no. 1, p. 32635, 2025.
[32] A. Tomashin, I. Gordon, and S. Wallot, "Interpersonal physiological synchrony predicts group cohesion," *Frontiers in human neuroscience,* vol. 16, p. 903407, 2022.
[33] Y. Liu, T. Wang, K. Wang, and Y. Zhang, "Collaborative learning quality classification through physiological synchrony recorded by wearable biosensors," *Frontiers in Psychology,* vol. 12, p. 674369, 2021.
[34] M. Dindar, S. Järvelä, A. Nguyen, E. Haataja, and A. Çini, "Detecting shared physiological arousal events in collaborative problem solving," *Contemporary Educational Psychology,* vol. 69, p. 102050, 2022.
[35] M. Worsley, "Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors," in *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, 2014, pp. 1-4.
[36] S. Mu, M. Cui, and X. Huang, "Multimodal data fusion in learning analytics: A systematic review," *Sensors,* vol. 20, no. 23, p. 6856, 2020.
[37] M. M. T. Rodrigo, E. A. Anglo, J. Sugay, and R. Baker, "Use of unsupervised clustering to characterize learner behaviors and affective states while using an intelligent tutoring system," in *International Conference on Computers in Education*, 2008, pp. 57-64.
[38] W. Xu, Y. Wu, and F. Ouyang, "Multimodal learning analytics of collaborative patterns during pair programming in higher education," *International Journal of Educational Technology in Higher Education,* vol. 20, no. 1, p. 8, 2023.
[39] H. Acosta *et al.*, "Multimodal learning analytics for predicting student collaboration satisfaction in collaborative game-based learning," 2024.
[40] T. E. Johnson, Y. Lee, M. Lee, D. L. O'Connor, M. K. Khalil, and X. Huang, "Measuring sharedness of team-related knowledge: Design and validation of a shared mental model instrument," *Human Resource Development International,* vol. 10, no. 4, pp. 437-454, 2007.
[41] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal,* vol. 1, no. 6, pp. 90-95, 2013.
[42] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method," *Accounting, Auditing and Finance,* vol. 1, no. 1, pp. 5-8, 2020.
[43] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics,* vol. 20, pp. 53-65, 1987.
[44] A. K. Singh, S. Mittal, P. Malhotra, and Y. V. Srivastava, "Clustering evaluation by davies-bouldin index (dbi) in cereal data using k-means," in *2020 Fourth international conference on computing methodologies and communication (ICCMC)*, 2020: IEEE, pp. 306-310.
[45] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 569, no. 5: IOP Publishing, p. 052024.