An extended community of inquiry framework for monitoring and predicting online peer learning participation

Journal of Applied Research in Higher Education

Received 17 April 2026
Revised 16 June 2026
Accepted 2 July 2026

Mohsen Dokhanchi, Shahrzad Saremi, Rania Shibl, Maryam Heidari, Hassan Ahmed, Dahlia Mansoor, Yassine Himeur, Mohammad Al-Zaffin, Shadi Atalla and Wathiq Mansoor
(Information about the authors can be found at the end of this article.)

Abstract

Purpose – Online learning communities on social media platforms can support peer learning, but educators often lack theoretically grounded and measurable approaches for monitoring how participation and discourse evolve across a semester. This study proposes an extended Community of Inquiry (CoI) evaluation framework that integrates Social, Teaching, and Cognitive Presence with a fourth behavioural dimension, Student Presence.

Design/methodology/approach – A sequential exploratory mixed-method design was adopted. Qualitative analysis of prior literature and semester-long observations of two large first-year engineering course Facebook groups (each enrolling 800–1000 students) informed an indicator-based coding scheme, applied quantitatively over Weeks 1–13. Predictive modelling used a persistence baseline, a multi-output Random Forest, and a multilayer perceptron under time-aware evaluation protocols.

Findings – Social Presence was enquiry-driven and peaked in Weeks 3–4; Teaching Presence was frontloaded and primarily reactive; Cognitive Presence was shallow, dominated by remembering and analysing. Student participation was consumption-oriented, with observers consistently outnumbering posters. Random Forest achieved consistent poster prediction (R^2 ~ 0.48–0.49), while observers and non-members remained difficult to forecast due to structural interdependence. Permutation importance identified remembering and evaluating as the most influential cognitive predictors.

Research limitations/implications – The dataset comprises 13 weekly observations from a single platform and institution, limiting generalisability. Future work should collect multi-cohort data, introduce lagged predictors, and explore individual-level modelling.

Practical implications – The framework provides instructors with an early-warning system for low poster activity, enabling timely, evidence-based interventions to support online peer learning communities.

Originality/value – This study makes three contributions: a multi-dimensional coding scheme grounded in the extended CoI framework; a data-driven analytics pipeline enabling descriptive monitoring and predictive modelling of participation roles; and an integrated evaluation framework that combines theory-grounded indicator coding with transparent machine learning to produce actionable insights from social media learning data.

Keywords Community of inquiry, Online peer learning, Student presence, Predictive modelling, Learning analytics

Paper type Research article

1. Introduction

Digital technology has fundamentally transformed how students learn and collaborate. Online learning communities are now integral to higher education, offering flexible environments that extend classroom interaction beyond physical boundaries (Bruggeman *et al.*, 2021). Within

Journal of Applied Research in Higher Education
Vol. 18 No. 8, 2026
pp. 113-141
Emerald Publishing Limited
e-ISSN: 1758-1184
p-ISSN: 2050-7003
DOI 10.1108/JARHE-04-2026-0666

these environments, peer interaction is not incidental — it is foundational. Research consistently shows that both instructor–learner and peer-to-peer exchanges are essential for building knowledge, sustaining engagement, and fostering a sense of belonging (Freeman and Jarvie-Eggart, 2019; Garrison, 2022; Lai *et al.*, 2019).

As online education expands, understanding what makes peer interactions effective has become a pressing concern. However, evaluating the quality and impact of these interactions in digital settings remains an open challenge (Babik *et al.*, 2024). Most existing approaches assess isolated dimensions, such as cognitive gains, while privileging outcomes over the processes through which learning unfolds (Yu and Schunn, 2023). Peer learning is inherently dynamic. It evolves in response to shifting group dynamics, course milestones, and individual participation patterns — none of which static, outcome-focused measures adequately capture (Chen *et al.*, 2024; Moon *et al.*, 2024; Zhang *et al.*, 2025). A unified, multi-dimensional framework that evaluates both the process and progress of peer learning over time is therefore lacking.

To address this gap, this study investigates effective methods for assessing peer learning in online learning environments. A mixed-methods investigation was conducted at an Australian university, where student and instructor interactions were analysed across a full semester. The study context comprised two compulsory first-year engineering courses at The University of Queensland, using Facebook-based course discussion groups as the online platform. This setting provided a rich, naturalistic dataset spanning thirteen weeks, enabling analysis of both the processes and outcomes of student participation.

The proposed solution is a comprehensive evaluation framework with two integrated components. The first is a theory-grounded coding scheme informed by the Community of Inquiry (CoI) model (Garrison *et al.*, 1999) and extended through the inclusion of student presence (Dokhanchi *et al.*, 2018). It operationalises four dimensions of peer learning discourse — student, cognitive, teaching, and social presences — into measurable weekly indicators. The second is a data-driven analytics pipeline that couples with the coding scheme to support descriptive monitoring of engagement patterns and exploratory prediction of student participation roles over time. Together, these components equip instructors and researchers with a practical tool for tracking how online peer learning communities develop and identifying which discourse patterns are associated with active, meaningful participation.

The contributions of this study are as follows:

(1) A multi-dimensional coding scheme grounded in the CoI framework and extended with student presence, operationalising four dimensions of peer learning discourse into measurable weekly indicators for systematic evaluation.

(2) A data-driven analytics pipeline coupled with the coding scheme to enable descriptive monitoring of weekly engagement patterns and predictive modelling of student participation roles over time.

(3) A practical, integrated evaluation framework that empowers instructors and researchers to analyse learner–learner and learner–instructor interactions, inform instructional design, and identify opportunities for improving online learning communities.

The remainder of this paper is organised as follows. Section 2 reviews related work. Section 3 details the methodology. Section 4 presents the experiments and results. Section 5 discusses limitations and future directions. Section 6 concludes the paper.

2. Related work

Researchers have previously highlighted the importance of interaction in online learning environments (Freeman and Jarvie-Eggart, 2019; Lai *et al.*, 2019). Collaboration emerging from learner–learner and learner–instructor interactions is the key contributor to successful learning (Palloff and Pratt, 1999), as such interactions help learners test ideas, receive feedback, and refine their understanding (Wagner, 1997). While interactions positively influence students' satisfaction and engagement (Liu *et al.*, 2017), they must be structured and cohesive to enable higher levels of critical thinking and knowledge construction (Garrison and Cleveland-Innes, 2005).

Multiple scholars have attempted to categorise types of interaction. For example, Wagner (1997) focused on learners' achievement, whereas Hare *et al.* (1994) distinguished between task-driven and socio-emotional interactions. Task-driven interactions emphasise completing assigned work and tend to be instructor-oriented, while socio-emotional interaction focuses on relationships between learners (Rovai, 2002) and incorporates individual factors such as personality and emotions (Delahunty *et al.*, 2014).

There are four types of interactions in online learning: learner–learner, learner–instructor (Moore, 1989, 1993), learner–content (Liu and Kaye, 2016), and learner–interface (Hillman *et al.*, 1994). Table 1 summarises these four widely recognised types of interactions in online learning environments.

Given the focus of this study on peer learning, the review now turns to the two dimensions most relevant to peer learning: learner–learner and learner–instructor interactions.

2.1 Learner–learner interaction

Learner–learner interactions in online learning communities play a critical role in fostering a sense of community (Xie and Yen, 2011), supporting cognitive development (Lin *et al.*, 2017), and strengthening interpersonal skills (Lee, 2002). Although active contributors positively shape collaborative learning, a high proportion of passive participants or "lurkers" can negatively impact group interaction dynamics.

Several frameworks have been proposed to analyse learner–learner online interactions. Bravo *et al.* (2008) introduced a three-stage model consisting of observation, abstraction, and intervention, where interaction data are first observed and collected then analysed, and finally used to design improvements to collaboration. Similarly, Ke and Xie (2009) developed a framework based on Cercone (2008) deep learning model incorporating social interaction, knowledge construction, and regulation of learning. Vuopala *et al.* (2016) created a coding scheme distinguishing between task-related and group-related interactions to assess collaborative learning processes.

These frameworks highlight the multidimensional nature of learner–learner interaction, reinforcing the need for evaluation models that capture both social and cognitive aspects of peer learning.

2.2 Learner–instructor interaction

Instructor involvement is essential in online peer learning communities because peer interaction alone does not always lead to productive outcomes (Kanuka and Garrison, 2004;

Table 1. Types of interaction in online learning environments

Interaction type	Definition	Citation
Learner–learner	Interaction between one learner and other learners, with or without an instructor present	Moore (1989)
Learner–instructor	Interaction between the learner and an expert who prepared or delivers course material	Moore (1989)
Learner–content	Cognitive or perceptual contact between students and study materials that results in the acquisition of meaning	Liu and Kaye (2016)
Learner–interface	The process of manipulating tools to accomplish a learning task	Hillman *et al.* (1994)

Xie *et al.*, 2018; Zhu, 2006). Instructors structure discussions, guide inquiry, and sustain engagement through feedback (Shute, 2008). Majeski and Stover (2007) show that instructor facilitation shapes the depth of discourse, moving students from surface-level exchange toward higher-order reflection. This study therefore analyses instructor and student interactions in parallel rather than treating them as separate phenomena.

2.3 Dimensions for evaluating interactions in online learning communities

This study adopts the extended Community of Inquiry (CoI) framework, which provides a multidimensional lens for analysing interactions in online learning environments. Garrison *et al.* (1999) originally identified three interrelated presences—cognitive, social, and teaching—as overlapping processes that shape the quality of the learning experience and determine student engagement in online education. Cognitive presence reflects the extent to which learners develop meaningful understanding through inquiry and reflection. Social presence refers to the learners' ability to present themselves authentically and build connections within the learning community. Teaching presence encompasses the design, facilitation, and instructional direction provided by the instructor. Later extensions to the model introduced Student Presence, which emphasises learners' self-regulation, agency, and active participation in managing their learning (Dokhanchi *et al.*, 2018). A study conducted by Dokhanchi *et al.* (2018) found that student presence, cognitive presence, teaching presence, and social presence affect peer learning in online learning environments. These four dimensions collectively provide a holistic framework for evaluating the complexity of interactions within online peer learning environments.

2.3.1 Cognitive presence. Garrison *et al.* (1999) categorised students' cognitive engagements as triggering events, exploration, integration, and resolution. Building on this work, Zhu (2006) examined student cognitive engagement processes in online learning communities, identifying processes such as seeking, interpreting, analysing, summarising information, and making decisions. More recently, Fiock (Fiock (2020) provided a comprehensive review of strategies to enhance cognitive presence in online learning environments, aligning instructional design with the Community of Inquiry framework to support deeper engagement and critical thinking.

2.3.2 Social presence. To measure social presence in online learning communities, Rourke *et al.* (1999) coded students' messages into affective, interactive, and cohesive categories. Swan and Shih (2005) expanded these categories by adding indicators such as expressing values and social expression. Similarly, Vuopala *et al.* (2016) analysed students' social interactions by including expressing cohesion, decreasing tension, and accompanying indicators in their coding scheme. Borup *et al.* (2012) used emotional expressions, open communication, and cohesion categories to analyse social presence. According to Borup *et al.* (2012), the presence of group cohesiveness indicators in online learning communities indicates the existence of a sense of community among members.

2.3.3 Teaching presence. Teaching presence has been conceptualised in various ways. Berge (1995) coded instructor posts into managerial, social, pedagogical, and technical indicators, while Anderson *et al.* (2001) organised teachers' presence around instructional design and organisation, facilitating discourse, and direct instruction. More recently, Stenbom (Stenbom (2018) reviewed empirical research on teaching presence in online environments and proposed a refined model highlighting the dynamic and context-dependent nature of instructional roles. Teaching presence is particularly relevant in peer learning contexts, as instructors play a crucial role in structuring discussions, guiding inquiry, and sustaining student engagement.

2.3.4 Student presence. Beyond the original three dimensions of the CoI framework, scholars have increasingly recognised the importance of a fourth dimension, often referred to as Student Presence. Early work by Shea and Bidjerano (2009) introduced the idea of Learner Presence as an extension of the model, emphasising the role of self-regulation and learner agency in online education. Their subsequent study (Shea and Bidjerano, 2012) further positioned Learning Presence as a moderating factor within the CoI framework and provided empirical evidence of its significance. Building on this perspective, Redmond (2014) highlighted the centrality of learner agency and self-directed effort in shaping meaningful engagement. More recently, studies identified Student Presence as the missing presence in the CoI framework, formally proposing it as a distinct and essential dimension (Ng *et al.*, 2021). Similarly, Dokhanchi *et al.* (2018) adopted Student Presence as the fourth presence, framing it as a reflection of students' active participation and responsibility within online learning communities.

Figure 1 shows the theoretical relationships among the four presences, and situates the extended CoI dimensions within the analytical pipeline of the proposed framework. End-to-end pipeline of the proposed framework is shown in Figure 2. The qualitative phase produces the coding scheme and indicator definitions. The quantitative phase aggregates coded indicators into weekly feature vectors, trains three competing models, and evaluates them under time-aware protocols before extracting interpretability results.

3. The proposed framework

To address the research question and meet the study objectives, a *mixed-method* research design was adopted (Creswell and Creswell, 2017). This approach was selected to capture both the interpretive depth of peer-learning discourse and the measurable breadth of behavioural participation in online learning environments. Importantly, the choice of mixed methods aligns with the multi-dimensional nature of the proposed evaluation framework, which integrates *Social Presence*, *Teaching Presence*, *Cognitive Presence*, and *Student Presence*. The study followed a *sequential exploratory* strategy in which qualitative insights were used to construct the coding scheme and quantitative analysis was then used to apply, validate, and summarise the resulting indicators at scale.

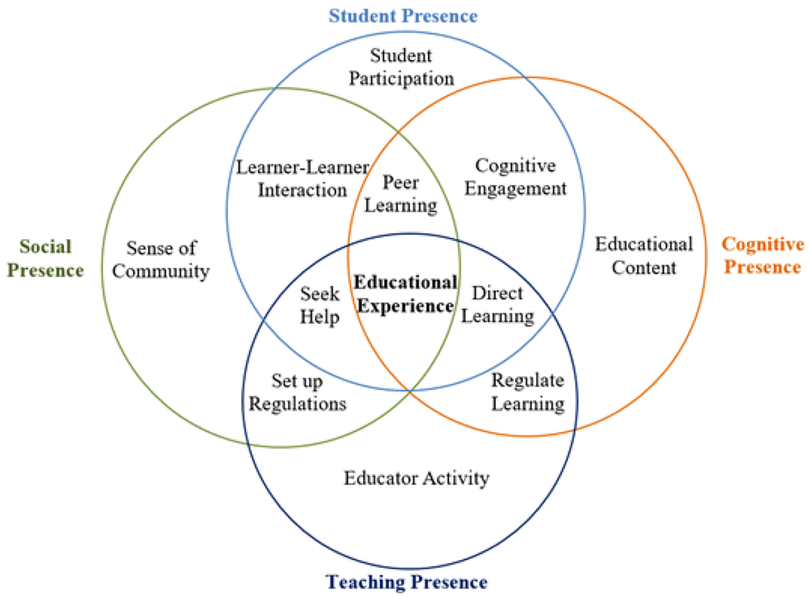

Figure 1. Extended CoI framework in online learning environments with student presence

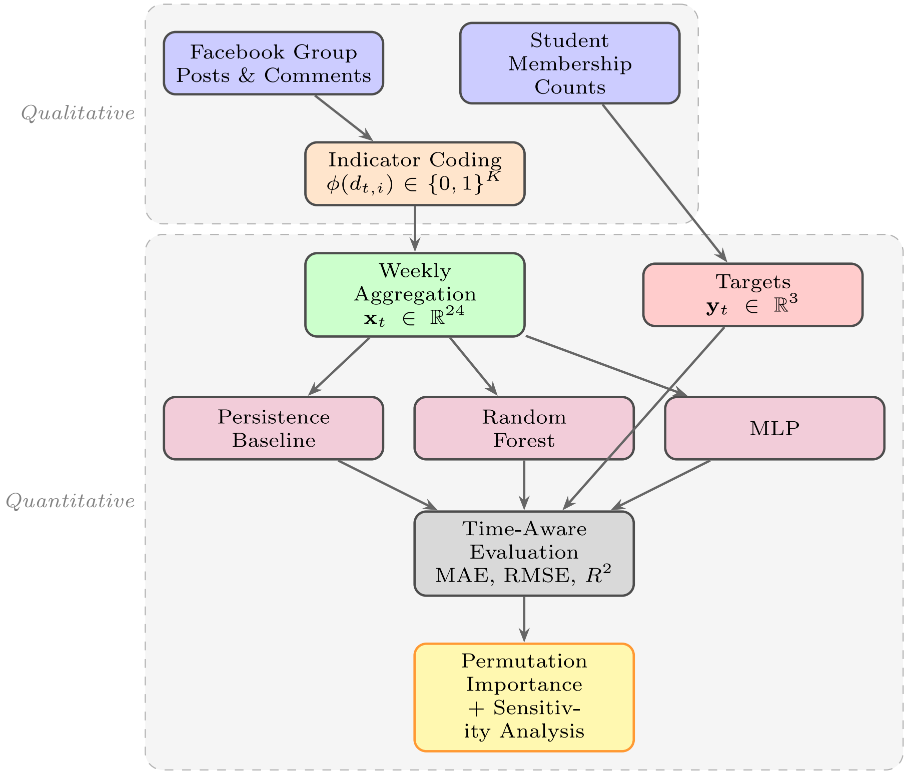

Figure 2. End-to-end pipeline of the proposed framework

In the qualitative phase, an extensive review of relevant literature was conducted alongside systematic observation of student and instructor interactions within the online learning communities across the semester. This stage provided an in-depth understanding of how peer learning unfolded over time and guided the iterative development of coding categories and indicators grounded in the extended CoI framework. In particular, qualitative analysis supported the refinement of indicator definitions, the clarification of boundary cases, and the alignment of the coding scheme with the communicative practices typical of social media environments. Building on this foundation, the quantitative strand operationalised the developed evaluation scheme to measure weekly interaction patterns. This enabled the research team to quantify participation levels, characterise the distribution of the four presences, and examine how these dimensions manifested and co-evolved across the semester as part of the peer-learning process.

The study was conducted at the University of Queensland using two compulsory first-year engineering courses. These large, team-based courses, each enrolling approximately 800–1000 students annually, were selected because they rely heavily on collaboration and peer interaction, making them well-suited for investigating online peer learning at scale. Course Facebook groups served as the primary online discussion platforms, offering an authentic, student-driven setting for observing peer support, instructor facilitation, and knowledge construction behaviours. Ethical approval for the study was obtained from the University of Queensland (Approval No. 2018000264), and permission to collect and analyse data was granted by the respective course coordinators.

3.1 Theoretical positioning of the extended CoI model

The proposed framework integrates (1) a theory-grounded coding scheme based on the extended Community of Inquiry (CoI) model (Garrison *et al.*, 1999) and (2) a data-driven modelling and validation layer to quantify, predict, and interpret student participation dynamics in online peer learning communities. The coding scheme was developed through an iterative process informed by the original CoI framework and its later extension incorporating student presence. The categories and indicators were designed to reflect the key factors that affect online peer learning—namely *student*, *cognitive*, *teaching*, and *social presences* (Dokhanchi *et al.*, 2018)—through: (1) reviewing existing coding schemes in the literature and (2) analysing student and instructor interactions in course Facebook groups across Weeks 1–13. Table 2 summarises the parameters used to guide the development of the framework.

Beyond descriptive analytics, the proposed framework contributes a modelling component that learns relationships between weekly aggregates of Social/Teaching/Cognitive indicators (inputs) and Student Presence participation roles (outputs). Specifically, the modelling layer aims to predict weekly Student Presence role counts (Poster, Observer, Non-Member) from weekly aggregates of Social, Teaching, and Cognitive presence indicators. This predictive component serves two purposes: (1) evaluation of whether the coded indicators contain sufficient signal to forecast participation behaviour and (2) insight generation by identifying which indicators are most strongly associated with student role changes over time.

To ensure methodological rigour for time-ordered data, the evaluation uses (1) a time-based holdout split (train on Weeks 1–10; test on Weeks 11–13) and (2) a walk-forward (rolling-origin) backtest that repeatedly trains on early weeks and predicts the next unseen week. These protocols reduce temporal leakage and provide a more credible estimate of performance under realistic deployment settings.

Table 3 summarises how each presence dimension was extended in this study relative to its original formulation.

3.2 Mathematical formulation of the weekly coding process

Let $t \in \{1, \ldots, T\}$ index academic weeks, where $T = 13$. Each week contains a set of discourse artefacts (posts and comments) denoted $\mathcal{D}_t = \{d_{t,1}, \ldots, d_{t,n_t}\}$, where n_t is the count of artefacts in week t. A coding function maps each artefact to a K-dimensional binary indicator vector:

$$\phi(d_{t,i}) \in \{0, 1\}^K, \tag{1}$$

Where $\phi_k(d_{t,i}) = 1$ if indicator k is present in artefact $d_{t,i}$ and 0 otherwise. A single artefact can receive multiple indicators (multi-label coding). Weekly counts are obtained by summing across artefacts:

$$x_{t,k} = \sum_{i=1}^{n_t} \phi_k(d_{t,i}), \qquad k = 1, \ldots, K. \tag{2}$$

The weekly feature vector concatenates all indicator counts:

Table 2. Parameters in online learning communities that were used to develop the framework

Dimension	Parameter
Student presence	Level of student participation in discussions
Cognitive presence	Students' course-related posts that include cognitive presence indicators
Social presence	Students' course and non-course related posts that include social presence indicators
Teaching presence	Course coordinators' and tutors' course and non-course related posts that include teaching presence indicators

Table 3. Theoretical contributions of each presence dimension relative to original CoI formulations

Dimension	Original formulation	Extension in this study
Cognitive presence	Triggering event, exploration, integration, resolution (Garrison *et al.*, 1999)	Reframed using Bloom's revised taxonomy (Anderson and Bloom, 2001): six cognitive processes from remembering through creating
Social presence	Affective, interactive, cohesive categories (Rourke *et al.*, 1999)	Extended with conversational expression (enquiry, statement, sharing information) and affective expression (paralanguage, self-disclosure)
Teaching presence	Instructional design, facilitating discourse, direct instruction (Anderson *et al.*, 2001)	Applied to instructor posts in an informal social media context; distinguishes reactive from proactive instructor behaviour
Student presence	Learner Presence emphasising self-regulation (Shea and Bidjerano, 2009)	Operationalised as observable weekly role counts: Poster, Observer, Non-Member (Dokhanchi *et al.*, 2018); used directly as prediction targets

$$\mathbf{x}_t = [x_{t,1},\ x_{t,2},\ \ldots,\ x_{t,K}]^\top \in \mathbb{R}^K. \tag{3}$$

In this study, x_t combines indicators from Social, Teaching, and Cognitive Presence, yielding $K = 24$ predictors.

3.3 Student presence as a multi-output response

Student Presence is represented by three weekly role counts:

$$\mathbf{y}_t = \left[y_t^{(P)},\ y_t^{(O)},\ y_t^{(N)}\right]^\top \in \mathbb{R}^3, \tag{4}$$

Where $y_t^{(P)}$, $y_t^{(O)}$, and $y_t^{(N)}$ denote Poster, Observer, and Non-Member counts respectively. The full dataset is then the set of paired weekly observations:

$$\mathcal{S} = \{(\mathbf{x}_t,\ \mathbf{y}_t)\}_{t=1}^{T}. \tag{5}$$

Because the three targets sum to total enrolment in each week (modulo rounding due to joining and leaving), they are structurally coupled. Specifically:

$$y_t^{(P)} + y_t^{(O)} + y_t^{(N)} \approx N_{\mathrm{enrol}}, \tag{6}$$

Where N_{enrol} denotes total enrolment. This linear dependence means that an accurate prediction of any two targets almost determines the third, but it also means that errors in one target propagate to the others. This structural constraint is an important consideration when interpreting multi-output model performance.

3.4 Predictive modelling objective

The modelling layer learns a function $f_\theta(\cdot)$ that maps presence indicators to weekly participation roles:

$$\widehat{\mathbf{y}}_t = f_\theta(\mathbf{x}_t), \tag{7}$$

Where θ denotes model parameters. Fitting is expressed as empirical risk minimisation:

$$\theta^{\star} = \text{argmin}_{\theta} \frac{1}{|\mathcal{T}_{\text{train}}|} \sum_{t \in \mathcal{T}_{\text{train}}} \|\mathbf{y}_t - f_{\theta}(\mathbf{x}_t)\|_2^2. \tag{8}$$

3.5 Baselines and models

Three approaches were used.

3.5.1 Persistence baseline. The simplest reasonable baseline predicts next week's counts using the current week's values:

$$\widehat{\mathbf{y}}_t^{(\text{pers})} = \mathbf{y}_{t-1}, \qquad t \geq 2. \tag{9}$$

A model that cannot outperform this baseline provides no practical forecasting value.

3.5.2 Random Forest (multi-output regression). Random Forest averages predictions from an ensemble of M decision trees:

$$\widehat{\mathbf{y}}_t^{(\text{RF})} = \frac{1}{M} \sum_{m=1}^{M} h_m(\mathbf{x}_t). \tag{10}$$

Random Forest is well suited to small tabular datasets because it captures non-linear feature interactions without requiring large training samples, and it produces interpretable importance scores as a by-product of training.

3.5.3 Multilayer perceptron (MLP). A deep multi-output regressor applies a composition of affine transformations and nonlinearities:

$$f_{\theta}(\mathbf{x}) = \mathbf{W}_3\, \sigma(\mathbf{W}_2\, \sigma(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3, \tag{11}$$

Where $\sigma(\cdot)$ is a nonlinear activation function (ReLU in this implementation) and $\theta = \{\mathbf{W}_\ell, \mathbf{b}_\ell\}_{\ell=1}^{3}$ are trainable parameters. The MLP was included to test whether a higher-capacity model improves performance relative to Random Forest when the training set is small.

3.6 Time-aware evaluation protocols

To prevent temporal leakage, two evaluation strategies were applied.

3.6.1 Time-based holdout split. The model was trained on Weeks 1–10 and tested on Weeks 11–13. This mirrors the situation where an instructor uses early-semester signals to forecast late-semester participation.

3.6.2 Walk-forward (rolling-origin) backtest. Given a minimum training size t_0, the model is retrained on weeks $\{1, \ldots, t-1\}$ and tested on week t for each $t = t_0 + 1, \ldots, T$:

$$\widehat{\mathbf{y}}_t = f_{\theta^{(t)}}(\mathbf{x}_t), \qquad \theta^{(t)} = \text{argmin}_{\theta} \frac{1}{t-1} \sum_{j=1}^{t-1} \|\mathbf{y}_j - f_{\theta}(\mathbf{x}_j)\|_2^2. \tag{12}$$

Walk-forward evaluation is more conservative than a single holdout split because each prediction is made strictly from past data, and the model must generalise across multiple time points rather than just three.

Figure 3 illustrates both evaluation protocols.

3.7 Performance metrics

For each target $q \in \{P, O, N\}$, mean absolute error (MAE) and root mean squared error (RMSE) measure average and worst-case prediction error:

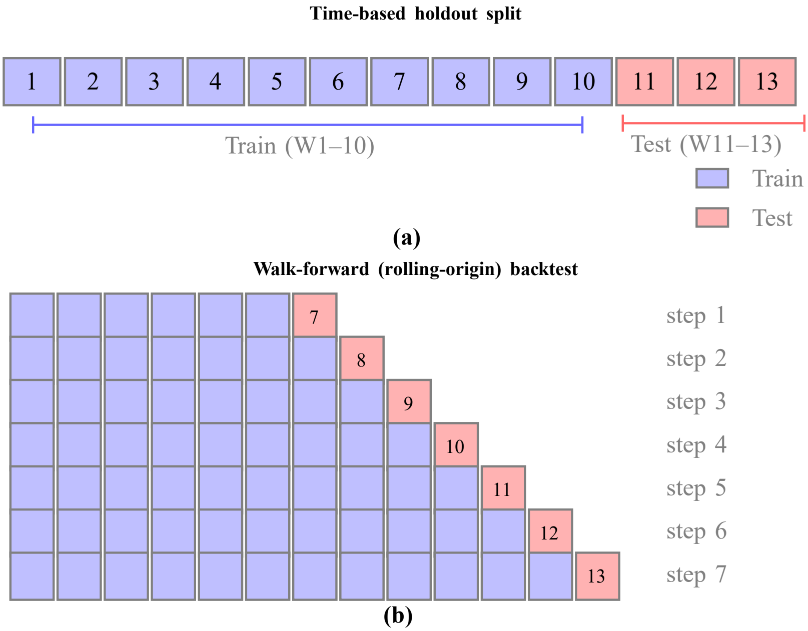

Figure 3. Evaluation protocols. Panel (a) shows the time-based holdout split: the model trains on Weeks 1–10 and is tested on Weeks 11–13. Panel (b) shows the walk-forward backtest: the model is retrained at each step on an expanding window and tested on the next single week, starting from a minimum training window of six weeks

$$\mathrm{MAE}^{(q)} = \frac{1}{|\mathcal{T}_{\text{test}}|}\sum_{t\in\mathcal{T}_{\text{test}}}\left|y_t^{(q)} - \widehat{y}_t^{(q)}\right|, \qquad \mathrm{RMSE}^{(q)} = \sqrt{\frac{1}{|\mathcal{T}_{\text{test}}|}\sum_{t\in\mathcal{T}_{\text{test}}}\left(y_t^{(q)} - \widehat{y}_t^{(q)}\right)^2}. \tag{13}$$

The coefficient of determination R^2 expresses explained variance relative to a mean baseline:

$$R^{2(q)} = 1 - \frac{\sum_{t\in\mathcal{T}_{\text{test}}}\left(y_t^{(q)} - \widehat{y}_t^{(q)}\right)^2}{\sum_{t\in\mathcal{T}_{\text{test}}}\left(y_t^{(q)} - \overline{y}^{(q)}\right)^2}, \tag{14}$$

Where $\overline{y}^{(q)}$ is the mean of the true values on the test interval. A negative R^2 indicates that the model performs worse than simply predicting the mean, which is itself a practically important finding in small-sample settings.

3.8 Interpretability: permutation importance and sensitivity analysis

To identify influential indicators, permutation importance is computed by measuring the degradation in predictive score after randomly permuting feature k:

$$\mathrm{PI}(k) = s(\mathbf{X}_{\text{test}}, \mathbf{Y}_{\text{test}}) - \mathbb{E}_\pi\left[s\left(\mathbf{X}_{\text{test}}^{\pi(k)}, \mathbf{Y}_{\text{test}}\right)\right], \tag{15}$$

Where $s(\cdot)$ is the model score (e.g. multi-output R^2), and $\mathbf{X}_{\text{test}}^{\pi(k)}$ denotes the test matrix with the k-th feature permuted.

Additionally, a one-dimensional what-if sensitivity analysis visualises the effect of varying feature k while holding all other features fixed at a reference vector x_{ref} (e.g. training mean):

$$g_k(v) = f_\theta\left(\mathbf{x}_{\mathrm{ref}} + \left(v - x_{\mathrm{ref},k}\right)\mathbf{e}_k\right), \quad (16)$$

Where e_k is the k-th canonical basis vector. This yields an interpretable curve relating feature values to predicted participation roles.

3.9 Coding scheme: presence indicators

3.9.1 Student presence. Student presence reflects the degree to which learners actively participate in the online community. Participation levels were observed and coded weekly across the semester, resulting in three categories: posters, observers, and non-members (Table 4). In addition to descriptive coding, these categories were operationalised as weekly target variables in the modelling layer, enabling the quantitative analysis of how Social/ Teaching/Cognitive presence indicators relate to participation dynamics.

Posters are active participants who create posts in online learning communities. Observers or lurkers do not actively participate in discussions; however, they learn by observing the discussions. They are passive participants and may become active participants at any time during the semester. Non-members in online communities can join the community anytime and become observers or posters during the semester.

3.9.2 Cognitive presence. In this research, the cognitive presence section of the coding scheme was developed based on observation of students' course-related posts adopting Bloom's taxonomy framework (Anderson and Bloom, 2001). Indicators reflect six cognitive processes: remembering, understanding, applying, analysing, evaluating, and creating. Table 5 presents the categories, definitions, and representative examples drawn from course Facebook group discussions.

In the modelling layer, cognitive presence indicators form a subset of the predictor vector x_t and quantify the cognitive depth of weekly discourse. This enables empirical testing of whether higher-order cognitive activity is associated with changes in participation roles, and supports interpretability analyses that identify which cognitive processes are most influential.

3.9.3 Social presence. Social presence captures how learners project themselves socially and emotionally within the online community. Students' course and non-course related posts in online discussions were observed. Table 6 presents the coding scheme that was developed in this research to analyse social presence in online learning communities. Students' posts and comments are categorised into group cohesiveness, conversational expression, and affective expression. Each category is divided further into different indicators.

In the modelling layer, social presence indicators represent interactional and affective signals and constitute a subset of x_t. This supports quantitative evaluation of whether socio-emotional projection and group cohesion co-vary with participation roles (Poster/Observer/ Non-Member), complementing the qualitative interpretation of discourse.

3.9.4 Teaching presence. Teaching presence was examined through course coordinators' and tutors' posts in online discussions. Indicators were organized into three categories: design

Table 4. Student presence indicators

Student type	Definition
Poster	A student who posts at least one message during the academic week
Observer	A student who has joined the community but does not post or comment. Students who like posts are included in this category
Non-member	A student who is enrolled in the course but has not yet joined the community

Table 5. Cognitive presence indicators

Category	Indicator	Student comment
Remembering	Recalling prior information	"Okay thank you, it's just that in the MEA guide it says to list them all so I just was a little confused."
Understanding	Confirming knowledge acquisition	"Thanks. It really helps! I can generate a simulation now!"
Applying	Carrying out a procedure	"For calibration. Simply set the Arduino to output a PWM signal of duty cycle 50%."
Analysing	Breaking knowledge into components to support understanding	"I was taking into consideration the dry weather with decreasing rainfall throughout the year where it's very cost effective."
Evaluating	Making a judgment based on available information	"Here is what we've got for the simulation result. The changes in colour seem alright in the image, but the value of the mass friction is quite close to 0.002 which is not that ideal."
Creating	Combining information components to produce new knowledge	"Considering the resolution of the camera is 1920x1080 pixels and knowing the distance it is looking at (say approximately 10 cm), you can calculate what the minimum strain you will be able to measure using the OSS."

Table 6. Social presence indicators

Category	Indicator	Definition	Student comment
Group cohesiveness	Group Reference	Referring to community members as "we"	"Will we have a PIR for ENGG1200?"
	Tagging	Acknowledging other students in conversation	"John those numbers have come directly from the project C PowerPoint slides this week"
Conversational expression	Sharing Information	A post that informs other students	"For PVC we got a 0.23 K loss for 3 L/s and 313K"
	Statement	An expression of views or ideas	"Everyone talks about pulling an all-nighter. No posts after 12 a.m."
	Enquiry	A question directed at other students	"Has anyone know whereabouts of a grey little pencil case that I left in the advanced engineering building today"
Affective expression	Emotional Icons	Use of icons to convey feelings	:)
	Emotional Words	Use of words to convey feelings	"You make me so happy. See you today"
	Paralanguage	Text forms used to convey feeling	"Whoa!!! Can't wait"
	Self-Disclosure	Expressing a personal feeling	"I am offended by this accusation"

and plan, facilitating discourse, and directing discourse. These categories capture instructors' roles in structuring learning activities, encouraging engagement, monitoring discussions, and providing clarification or guidance. Table 7 sets out the coding scheme developed to investigate teaching presence in online learning communities.

Teaching presence indicators are included in x_t as contextual signals reflecting instructional design and facilitation. Their inclusion enables the modelling layer to assess whether instructor behaviours (e.g. clarifying, giving directions, and encouraging) relate to transitions between participation roles over time.

Table 7. Teaching presence indicators

Categories	Indicator	Definition	Instructor comment
Design and plan	Setting Goals	Post that indicates regulations for participating in online discussions	"By signing up to this Facebook group you agree to use it responsibly."
	Defining Discussion Topics	Posts that direct discussions	"I'm giving an overview of the PIR in Wednesday's lecture. What would you particularly like to know?"
Facilitating discourse	Icebreaking	Post that warms up the conversation between participants	"Yup, a good video! Don't forget to bring your popcorn and 3D glasses for the ultimate movie experience!"
	Encouraging	Post that encourages students to participate in the discussion	"The first speakers for 2015! Well done on addressing 600 people"
	Monitoring	Comments that show the online discussions are being monitored	"Woop! Rule breakers identified! Request instructions on appropriate punishment"
	Acknowledging	Tagging students in the discussions	"Seriously Alex, we will make sure that he doesn't mark your logbook."
Directing discourse	Clarifying	Post that clarifies students' enquiries or statements	"Absolutely! Logbooks are a fantastic way to keep all your thoughts in a safe collected space!"
	Giving Directions	Post that directs students to other resources	"Full scale - it's in Doc 2 and my LCA slides."
	Asking Questions	Post that requires students' responses	"Which one are you talking about?"
	Providing Information	Post that provides information	"Google EAIT cover sheets, and all will be revealed"

4. Discussion and experimental results

Based on the weekly interaction analysis (Table 8) and the predictive modelling outputs, this section synthesises (1) descriptive insights about Social, Teaching, Cognitive, and Student Presence dynamics over Weeks 1–13 and (2) quantitative evidence from exploratory analysis, forecasting experiments, and interpretability methods. Together, these results provide a multi-perspective understanding of online peer learning behaviour in the course Facebook group.

4.1 Dataset

The merged dataset contains 13 weekly observations and 27 variables: 24 predictors (Social, Teaching, and Cognitive Presence indicators) and three targets (Poster, Observer, Non-Member). The small sample size has two practical consequences. First, any individual performance metric is sensitive to a single anomalous week, so MAE, RMSE, and R^2 must be interpreted cautiously and in combination. Second, complex models with many parameters (such as the MLP) are highly prone to overfitting. Because the dataset is small and time-ordered, model estimates and generalisation performance are highly sensitive to overfitting and to the chosen evaluation protocol. Therefore, the analysis combines descriptive statistics (Table 8), time-series visualisation, time-aware splits, walk-forward validation, and interpretability tools.

4.2 Presence dynamics across the semester

4.2.1 Social presence. Social presence recorded robust interaction over the semester (Table 8), particularly through Enquiry, which peaks in Weeks 3 and 4. This pattern suggests alignment with demanding course content and/or assessment pressure points that triggered clarification-

Table 8. Weekly interaction data: indicator counts per Week (W1–W13) for all four presence dimensions

Dim	Indicator	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13
Social	Group Ref	23	18	24	23	12	6	6	10	16	6	4	16	18
	Tagging	3	3	31	36	18	11	13	9	12	11	11	17	36
	Sharing Info	10	12	19	31	8	7	3	3	1	1	0	0	2
	Statement	30	39	11	39	19	13	10	6	7	7	6	5	7
	Enquiry	30	41	77	69	21	15	9	17	22	20	13	24	60
	Emot. Icons	20	15	26	27	9	7	6	8	9	8	4	6	17
	Emot. Words	4	5	10	9	5	1	2	1	1	2	1	3	5
	Paralanguage	3	6	22	27	6	3	1	3	8	5	3	4	13
	Self-Disclosure	3	5	8	10	3	2	1	1	3	1	1	0	2
Teaching	Setting Goals	2	0	0	0	0	0	0	0	0	0	0	0	0
	Defining Topics	3	0	0	0	0	0	0	0	0	0	0	0	0
	Icebreaking	35	32	46	38	7	19	9	13	7	4	6	3	45
	Encouraging	12	18	21	11	11	7	9	3	5	4	3	5	19
	Monitoring	0	1	3	2	0	0	1	0	1	0	0	0	0
	Acknowledging	22	17	31	32	3	10	1	4	4	4	4	2	8
	Clarifying	12	9	24	11	2	6	1	4	2	2	5	1	10
	Giving Dir	4	6	13	5	2	2	3	3	2	1	2	1	2
	Asking Ques	10	7	13	8	1	6	3	2	1	1	2	1	6
	Providing Info	34	55	80	76	13	22	16	18	20	18	15	4	28
Cognitive	Remembering	4	12	14	11	4	3	3	6	11	3	4	13	9
	Understanding	4	3	16	14	2	4	1	2	5	1	2	5	6
	Applying	0	1	6	4	1	0	1	0	1	1	1	2	0
	Analysing	6	15	37	29	6	4	4	12	18	7	5	8	4
	Evaluating	0	3	10	8	4	2	2	5	12	0	4	3	5
	Creating	0	1	5	4	2	2	1	2	2	0	1	4	1
Student	Poster	40	44	80	92	37	41	38	35	39	43	33	38	63
	Observer	360	416	440	450	514	513	520	523	519	516	528	526	507
	Non-Member	590	530	470	448	439	436	432	432	432	431	429	426	420

seeking and peer support. Sustained Statement and Tagging activity indicates an active conversational environment that supported informal collaboration. Emotional Icons and Emotional Words appear at moderate levels, while Self-Disclosure remains low, which is typical for formal academic contexts. Overall, the social presence evidence suggests that peer support for academic problem solving was prioritised over emotional bonding.

The dominance of Enquiry over Emotional Words and Self-Disclosure suggests that students used the Facebook group primarily for academic problem solving rather than social bonding. This is an important distinction for instructors: high Social Presence in an engineering course Facebook group does not necessarily imply community warmth; it may simply reflect high information-seeking activity around deadlines.

4.2.2 Teaching presence. Teaching Presence was highest in early weeks and concentrated in two indicators: Providing Information and Icebreaking. Setting Goals and Defining Discussion Topics occurred only in Week 1, and Monitoring was almost entirely absent. This pattern points to a reactive instructor role: instructors respond to student questions, but they do not proactively shape the discourse beyond establishing initial community norms.

A proactive Teaching Presence, in which instructors set discussion topics, prompt students with higher-order questions, and monitor discourse quality over time, may be a realistic lever for shifting Cognitive Presence from lower-order to higher-order processes. The data here do not allow a causal test of that hypothesis, but the pattern creates a testable implication for future studies.

4.2.3 Cognitive presence. Cognitive Presence was the least frequent dimension, and the discourse that was coded remained concentrated at the lower end of Bloom's taxonomy. Remembering and analysing together account for the majority of coded cognitive activity. Evaluating, Creating, and Applying each appeared infrequently, with weekly counts rarely exceeding five.

This finding is consistent with the informal nature of a social media discussion platform, where structured prompts that would support higher-order thinking (e.g. reflective tasks, case analyses) are absent. It also aligns with prior work showing that deeper cognitive engagement in online environments requires sustained instructional scaffolding (Fiock, 2020; Garrison and Cleveland-Innes, 2005).

4.2.4 Student presence. Observer counts dominated throughout the semester, ranging from 360 in Week 1 to a peak of 528 in Week 11. Poster counts were modest, peaking at 92 in Week 4 and then declining to a trough of 33 in Week 11. Non-Member counts declined monotonically from 590 in Week 1–420 in Week 13, indicating steady uptake of group membership. However, membership growth did not translate into posting: of the 170 students who joined between Weeks 1 and 13, virtually all became Observers rather than Posters.

The Week 4 Poster spike coincides with a peak in Enquiry (69), Analysing (29), Tagging (36), and Providing Information (76), suggesting that assessment pressure simultaneously raised the volume of discourse and the proportion of students who posted at least once. The Week 13 Poster spike (63) shows a similar but smaller pattern.

Figure 4 presents a schematic summary of presence dynamics across the semester, and Figure 5 shows the distribution of Cognitive Presence across Bloom's taxonomy levels.

4.3 Exploratory data analysis

4.3.1 Temporal Trends. Figures 6–7 visualise weekly dynamics of (1) the targets and (2) the presence indicators. Abrupt changes can reduce the effectiveness of simple baselines and hinder model generalisation, while smoother patterns can be exploited more reliably by machine learning models.

4.3.2 Target distributions. Figure 8 presents the empirical distribution of the targets. Differences in spread and skewness can explain discrepancies in predictive difficulty; highly variable targets are harder to forecast in small-sample settings.

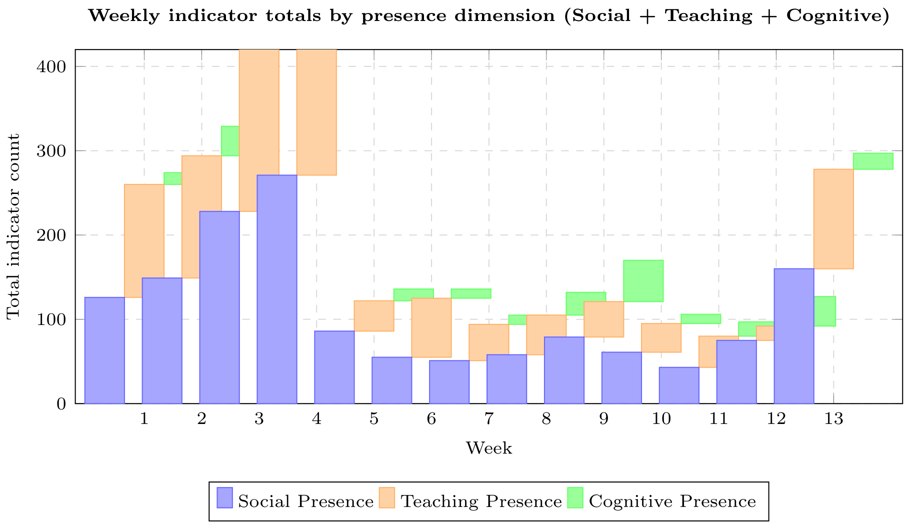

Figure 4. Weekly totals of social, teaching, and cognitive presence indicator counts. Activity peaks in Weeks 3–4, dips in Weeks 5–10, and partially recovers in Week 13. Cognitive Presence consistently contributes the smallest share of total discourse

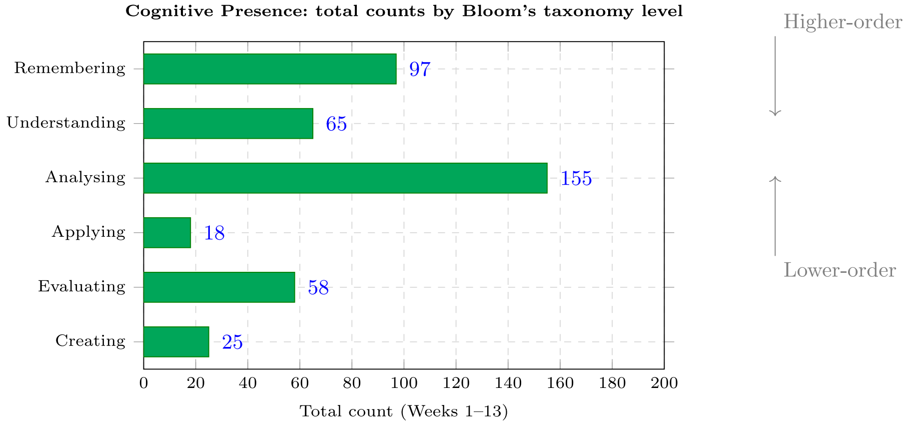

Figure 5. Distribution of cognitive presence coded indicators across Bloom's revised taxonomy levels (cumulative over Weeks 1–13). Analysing and Remembering dominate. Higher-order processes (Evaluating and Creating) appear at substantially lower rates, indicating that discourse rarely reached the level of knowledge synthesis or critical evaluation

4.3.3 Correlation structure. Figure 9 shows the correlation matrix over all variables, highlighting associations between presence indicators and student roles, predictor multicollinearity, and dependencies among targets.

(1) Top associations with targets (association only; not causal):

- Poster: Paralanguage (0.973), Enquiry (0.926), Understanding (0.915), Emotional Icons (0.873), Emotional Words (0.872), Self-Disclosure (0.862), Tagging (0.829), Sharing Information (0.826), Acknowledging (0.811), Analysing (0.765).

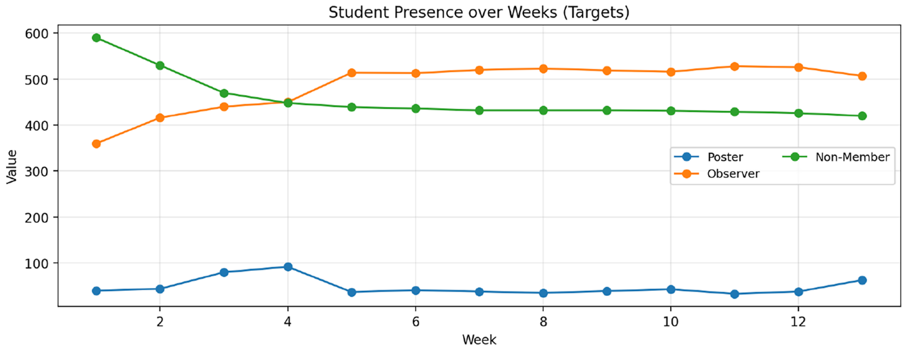

Figure 6. Student presence targets (poster, observer, non-member) over Weeks 1–13

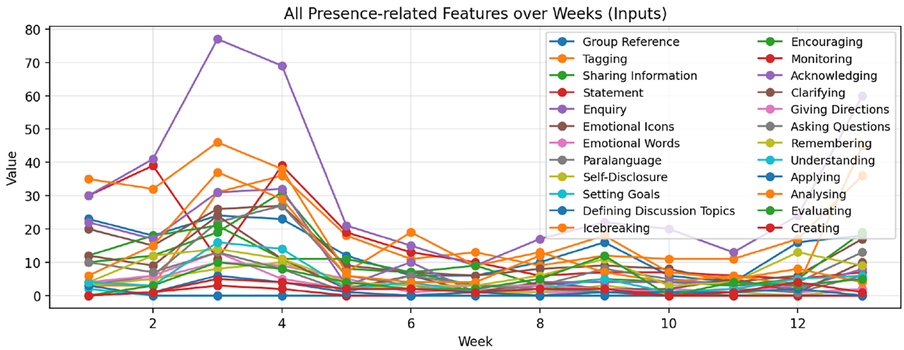

Figure 7. Presence-related input features over Weeks 1–13

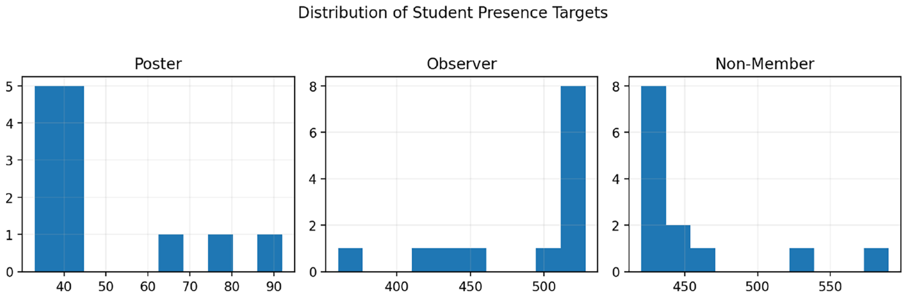

Figure 8. Histograms of the student presence targets

- Observer: Non-Member (0.937), Asking Questions (0.798), Acknowledging (0.782), Statement (0.767), Emotional Icons (0.738), Group Reference (0.727), Setting Goals (0.717), Defining Discussion Topics (0.717), Icebreaking (0.691), Clarifying (0.665).

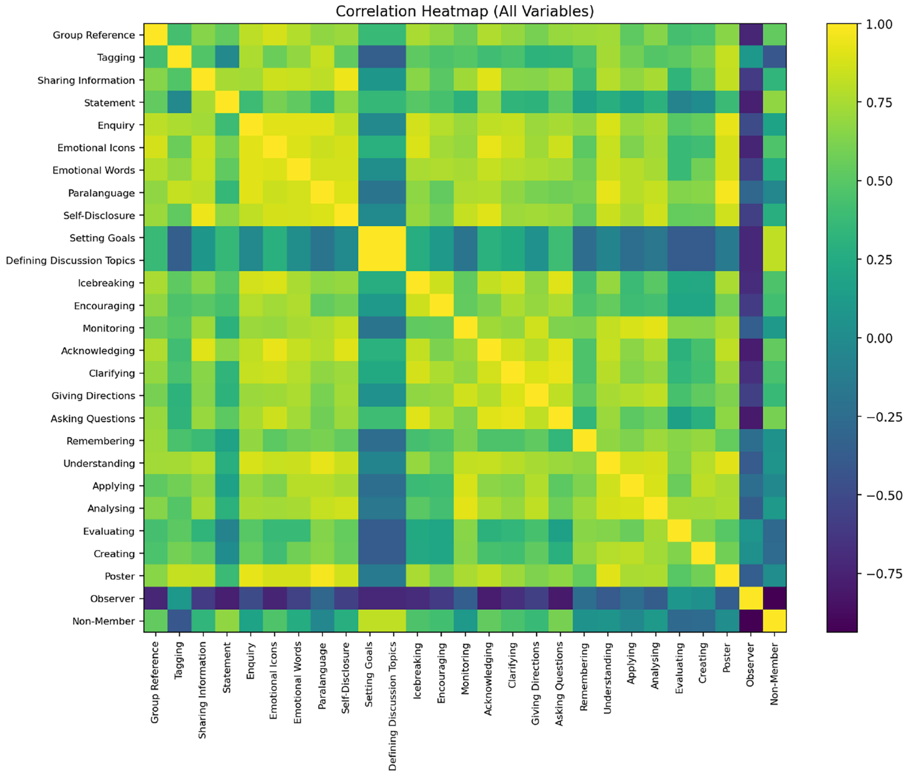

Figure 9. Correlation heatmap of all predictors and targets

- Non-Member: Observer (0.937), Setting Goals (0.816), Defining Discussion Topics (0.816), Statement (0.678), Asking Questions (0.605), Acknowledging (0.534), Group Reference (0.532), Emotional Icons (0.464), Icebreaking (0.458), Clarifying (0.444).

(2) Interpretation: Poster correlates strongly with socio-emotional signals (paralanguage, emotional cues, self-disclosure), suggesting that active posters exhibit richer social presence. Observer and Non-Member are strongly coupled ($r \approx 0.94$), indicating structural interdependence and complicating independent prediction.

4.4 Predictive modelling results

4.4.1 Evaluation protocol. A time-based split was adopted to reflect realistic forecasting conditions and prevent information leakage: models were trained on Weeks 1–10 and evaluated on unseen future weeks (Weeks 11–13). This setup mirrors how an instructor would predict upcoming participation from earlier signals. To strengthen reliability, a walk-forward (rolling-origin) backtest was also performed, repeatedly retraining on expanding histories and testing week-by-week.

4.4.2 Persistence baseline. The persistence baseline produced a negative R^2 for all three targets on the holdout set (Table 9). Poster showed the largest error (MAE = 13.33), reflecting the sharp jump in Week 13 (from 38 to 63) that the baseline could not anticipate. The negative R^2 values confirm that end-of-semester behaviour is not a simple continuation of mid-semester behaviour.

Table 9. Holdout performance (Weeks 11–13): persistence baseline

Target	MAE	RMSE	R^2
Poster	13.33	15.81	−0.45
Observer	11.00	13.03	−0.89
Non-member	3.67	4.04	−0.17
OVERALL	*9.33*	*12.06*	−0.58

4.4.3 Random Forest.

(1) Holdout evaluation (Weeks 11–13): To assess out-of-sample generalisation in a realistic forecasting setting, the RandomForest model was trained on the first ten weeks (Weeks 1–10) and evaluated on the final three weeks (Weeks 11–13), ensuring that no future information leaked into the training stage. Table 10 reports the resulting error metrics (MAE and RMSE) together with R^2 for each student-presence target and for the overall multi-output prediction. In addition to the tabulated metrics, Figures 10–11 provide complementary diagnostic views. Figure 10 plots the actual and predicted trajectories across the holdout weeks, allowing visual inspection of whether the model captures the direction and magnitude of changes near the end of the

Table 10. Holdout performance (Weeks 11–13): RandomForest (multi-output regression)

Target	MAE	RMSE	R^2
Poster	8.11	9.36	*0.49*
Observer	19.39	20.94	−3.90
Non-Member	21.00	27.10	−51.46
OVERALL	*16.17*	*20.50*	*−3.57*

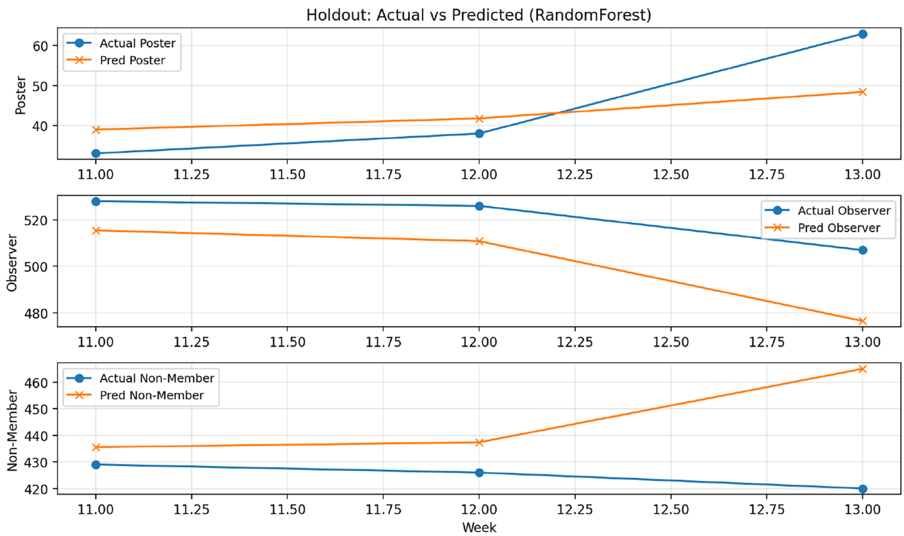

Figure 10. Holdout (Weeks 11–13): actual vs. predicted for RandomForest

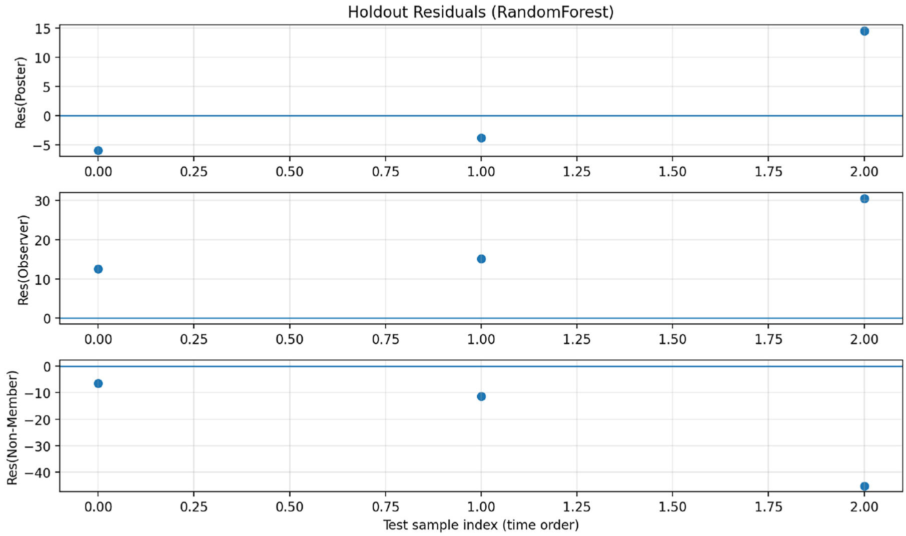

Figure 11. Holdout residuals (RandomForest)

semester. Figure 11 displays residual errors in time order, which helps identify systematic bias (consistent over- or under-estimation) and potential regime changes that may occur in the final weeks (e.g. assessment-related shifts). Together, these results quantify predictive accuracy while also revealing whether errors are stable or concentrated in specific weeks or targets, thereby supporting a more reliable interpretation of model performance under time-ordered educational data.

(2) Walk-forward backtest: To obtain a more robust estimate of time-series generalisation, a walk-forward (rolling-origin) backtest was performed. In this setting, the model is repeatedly retrained on an expanding window of past weeks and evaluated on the next unseen week, better reflecting deployment conditions. Table 11 summarises aggregate performance across all tested weeks, while Figure 12 visualises predicted versus observed values over time (see Table 12).

(3) Interpretation: RandomForest improves Poster prediction consistently, achieving positive R^2 values that indicate meaningful predictive signal in the presence of indicators for active participation. In contrast, Observer and Non-Member remain difficult to model, likely due to their strong interdependence (changes in one often imply changes in the other), temporal non-stationarity across weeks, and the limited sample size, which amplifies variance and sensitivity to outliers.

4.4.4 MLP. To examine whether a nonlinear deep model can capture more complex relationships among presence indicators, a multi-output MLP was trained on Weeks 1–10 and evaluated on the holdout period (Weeks 11–13). Table 12 reports performance metrics, while Figures 13–15 illustrate optimisation behaviour (training/validation loss) and residual errors, highlighting generalisation limitations under scarce data. The MLP performed poorly across all targets and evaluation settings (Table 12). The Observer R^2 of -751.39 and Non-Member R^2 of -2728.34 indicate catastrophic overfitting: the model fitted Week 1–10 patterns so tightly that it predicted values far outside the plausible range on the holdout. With only 10 training observations, a three-layer MLP with hundreds of parameters has far more capacity

Table 11. Walk-forward backtest: RandomForest performance

Target	MAE	RMSE	R^2
Poster	5.16	6.73	*0.48*
Observer	20.01	23.13	−11.70
Non-Member	20.55	25.61	−36.98
OVERALL	*15.24*	*20.30*	*−7.44*

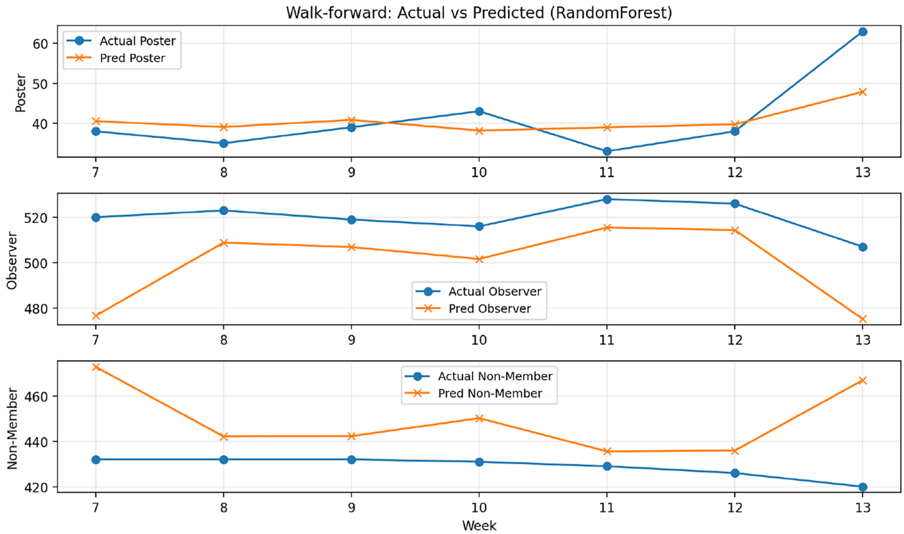

Figure 12. Walk-forward: actual vs. predicted for RandomForest

Table 12. Holdout performance (Weeks 11–13): MLP (multi-output regression)

Target	MAE	RMSE	R^2
Poster	25.45	30.33	−4.34
Observer	206.70	259.58	−751.39
Non-member	178.70	195.48	−2728.34
OVERALL	*136.95*	*188.42*	−385.22

than the data can support. This result confirms that deep models are not suitable for datasets of this size, regardless of their theoretical expressiveness.

Figure 16 provides a direct visual comparison of R^2 values for Poster across the three models and both evaluation protocols.

4.5 Interpretability

4.5.1 Feature importance. To better understand which indicators contribute most to prediction, feature relevance was examined using two complementary approaches. Impurity-based importance (Figure 17) provides a global view derived from the average reduction in split impurity across trees, but it can be biased under multicollinearity. Permutation importance

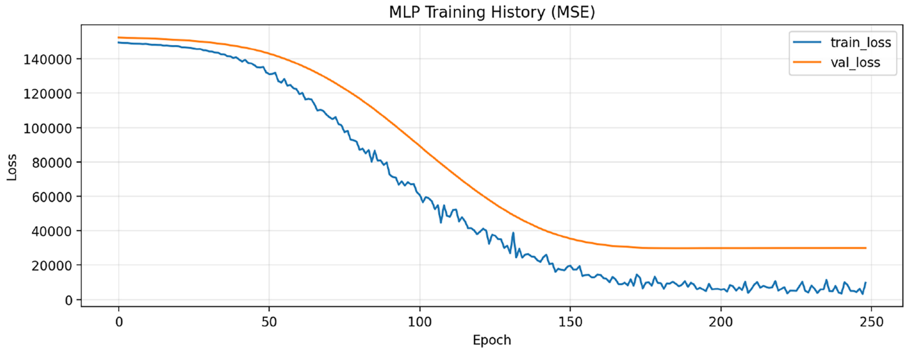

Figure 13. MLP training history (loss curves)

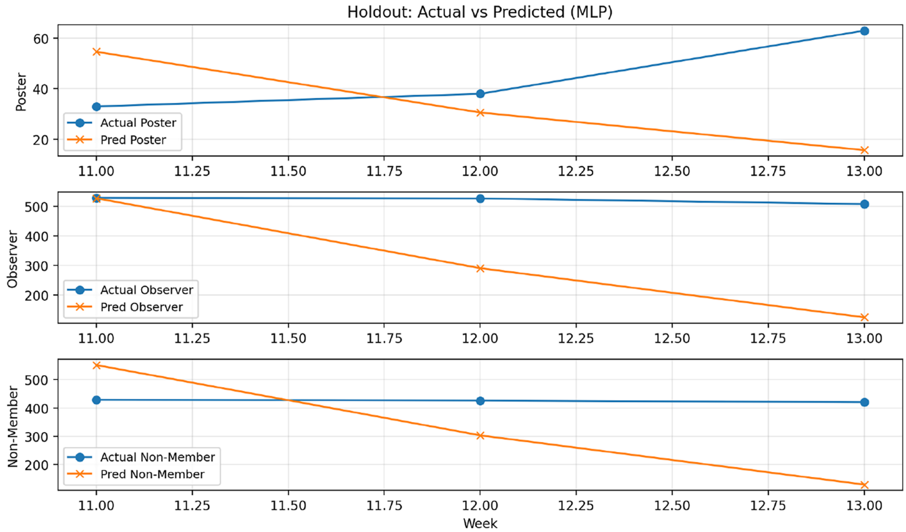

Figure 14. Holdout (Weeks 11–13): actual vs. predicted for MLP

(Figure 18) evaluates test-time influence by measuring performance degradation when a feature is randomly shuffled.

Acknowledging a Teaching Presence indicator appears next. Instructor posts that tag individual students may stimulate those students and their peers to become more active, although the correlational nature of this finding prevents a causal interpretation.

4.5.2 Sensitivity analysis. The sensitivity curves for Remembering show a non-linear relationship with Poster predictions: as Remembering counts increase from the 10th to the 90th percentile, predicted Poster counts rise and then plateau (Figures 19–20). A similar pattern appears for Evaluating. These curves suggest that moderate levels of cognitive discourse are associated with increases in posting, but gains taper off at high counts. This may reflect a ceiling effect: a class discussion can only sustain a limited number of active contributors regardless of how rich the cognitive content becomes.

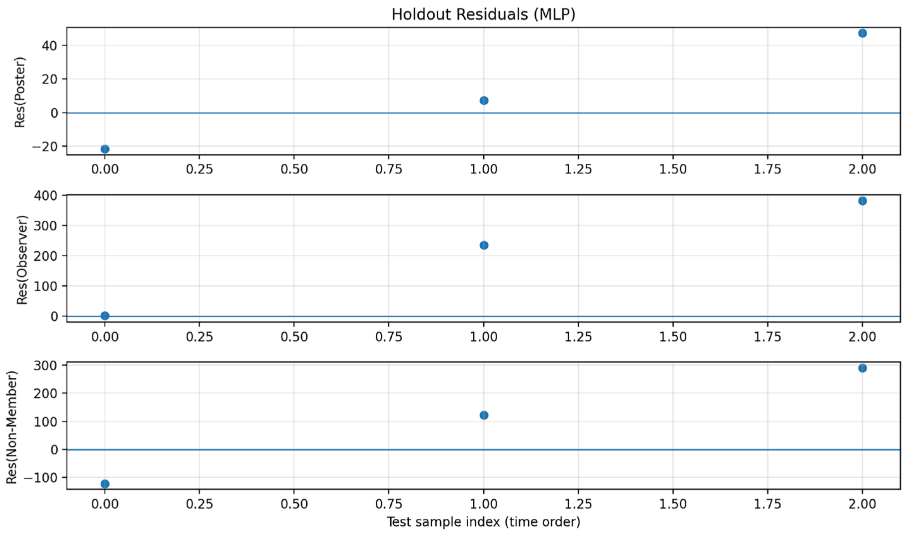

Figure 15. Holdout residuals (MLP)

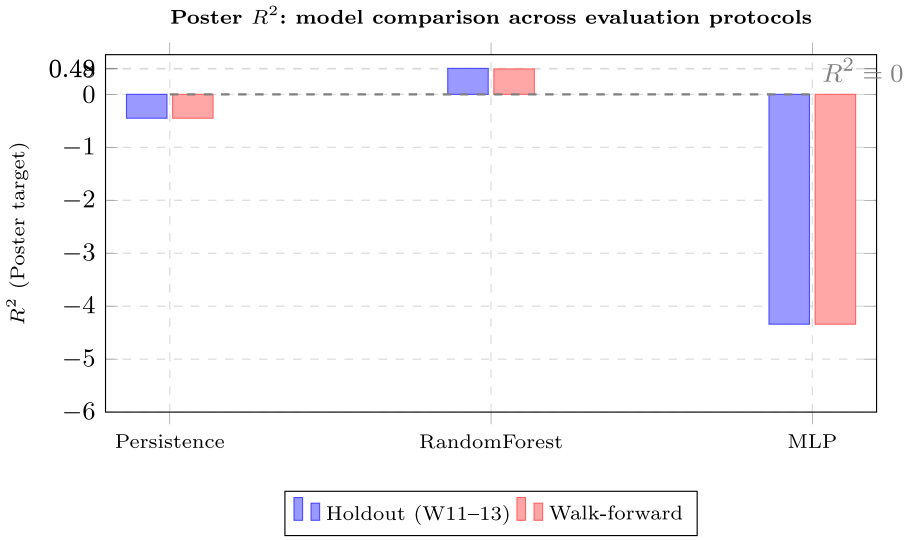

Figure 16. Comparison of poster R^2 values for the three models under holdout and walk-forward evaluation. positive R^2 values (above the dashed line) indicate that the model outperforms the mean baseline. Only Random Forest achieves this consistently. MLP and persistence fail on both protocols

4.6 Synthesis

Five findings emerge from the combined descriptive and predictive analysis.

(1) Social Presence is enquiry-driven and assessment-sensitive, peaking when deadlines approach and declining in low-pressure weeks.

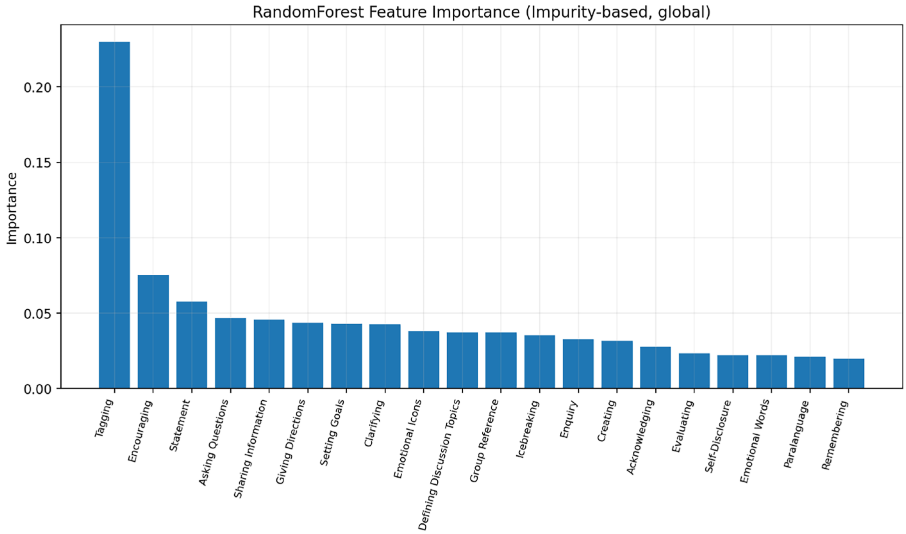

Figure 17. RandomForest feature importance (impurity-based; global)

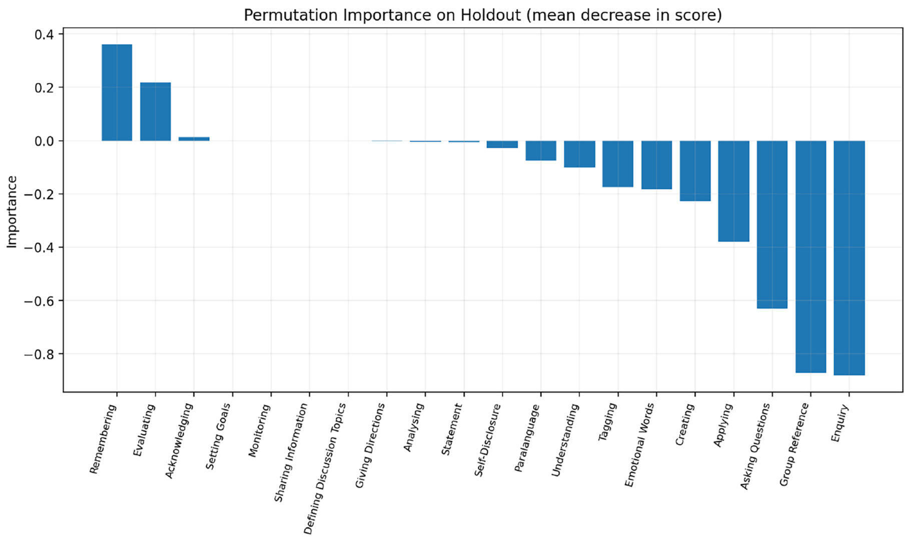

Figure 18. RandomForest permutation importance on the holdout set

(2) Teaching Presence is reactive and frontloaded: instructors provide information and break the ice in early weeks, but do not proactively structure discourse across the semester.

(3) Cognitive Presence is shallow, concentrated in remembering and analysing, with limited evidence of evaluating or creating.

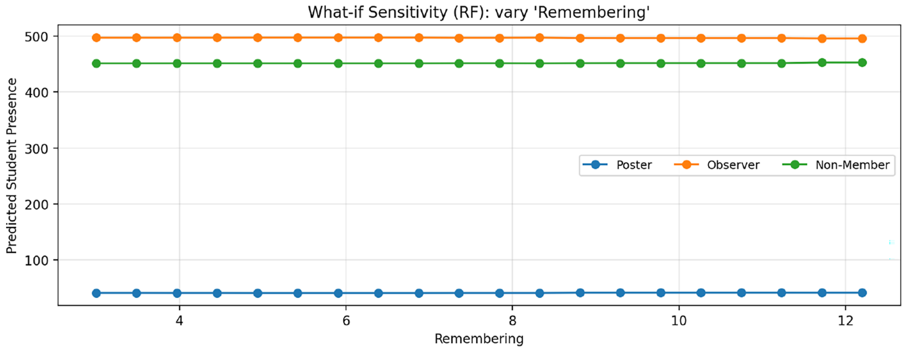

Figure 19. Sensitivity analysis: varying the top feature (remembering) while holding others fixed

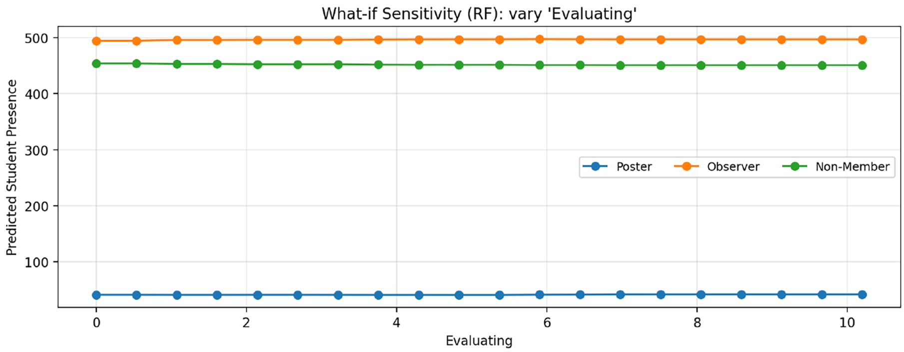

Figure 20. Sensitivity analysis: varying the second top feature (evaluating) while holding others fixed

(4) Student participation is consumption-heavy: the Observer role dominates throughout, and posting spikes selectively near assessment events rather than growing steadily.

(5) Random Forest predicts Poster behaviour with positive R^2 on both evaluation protocols, but Observer and Non-Member remain difficult due to structural coupling and slow within-semester change.

These findings converge on a practical implication: the presence indicators contain enough signal to predict active posting one to three weeks ahead with moderate accuracy, but they do not reliably predict passive participation levels. For instructors, this means the framework is most useful as an early-warning system for low Poster activity, not as a complete monitoring tool for all forms of participation.

5. Limitations and future directions

The study has several limitations that bound the generalisability and precision of its findings. The dataset contains only 13 weekly observations. At this scale, each R^2 value is estimated from three test points (holdout) or seven to eight points (walk-forward), making the metrics sensitive to individual anomalous weeks. Standard errors and confidence intervals on R^2 would require bootstrapping procedures that are not feasible with 13 total observations.

The data come from a single platform (Facebook), a single institution (The University of Queensland), and a single discipline (engineering). Facebook groups have specific structural and normative characteristics (e.g. visible likes, informal register, public posts by default) that may not generalise to other platforms such as discussion boards, Slack workspaces, or Microsoft Teams. Coding was performed by a single team, and inter-rater reliability statistics would strengthen confidence in the indicator definitions.

The weekly aggregation level obscures within-week dynamics. A student who posts on Monday of a high-Enquiry week and a student who posts on Friday contribute equally to the weekly Poster count, even though the causal mechanism linking the discourse to their posting decisions may differ.

Future work should address these limitations through several avenues. Collecting data from multiple cohorts and courses would provide the sample sizes needed for more stable estimates and cross-context comparisons. Introducing lagged variables (e.g. last week's Poster count as a predictor) would allow the model to exploit temporal autocorrelation explicitly. Modelling at the individual student level, where the target is a binary "did this student post this week" indicator, would enable a richer analysis and support targeted intervention design. Sequence models (e.g. LSTM or temporal convolutional networks) could be explored once larger longitudinal datasets are available.

6. Conclusion

This study proposed and tested an extended Community of Inquiry framework for monitoring and predicting online peer learning participation in a social media context. The framework integrates four dimensions: Social Presence, Teaching Presence, Cognitive Presence, and Student Presence, and couples a theory-grounded coding scheme with a data-driven analytics pipeline.

The coding scheme operationalises 24 indicators across three presence dimensions and defines three Student Presence roles (Poster, Observer, Non-Member) as observable weekly targets. Applied to two first-year engineering course Facebook groups over 13 weeks, the scheme produced a structured dataset that is small but theoretically rich. Descriptive analysis showed that Social Presence was enquiry-driven, Teaching Presence was reactive, Cognitive Presence was shallow, and student participation was consumption-oriented.

On the predictive side, Random Forest achieved a consistent Poster R^2 of approximately 0.48–0.49 on both holdout and walk-forward evaluation, outperforming the persistence baseline and confirming that the coded indicators carry predictive signal for active participation. Observer and Non-Member targets remained difficult to forecast, primarily because their structural interdependence and slow temporal dynamics leave little residual signal for a model to exploit. The MLP failed entirely under data scarcity, reinforcing the importance of selecting model capacity proportional to dataset size.

Permutation importance identified Remembering and Evaluating as the most influential predictors of posting behaviour. Although this finding is exploratory given the small holdout, it offers a plausible pedagogical interpretation: weeks in which students engage cognitively with course content, even at lower cognitive levels, tend to be weeks of higher active participation.

The framework demonstrates that theory-grounded indicator coding can be combined with transparent machine learning to produce actionable and interpretable insights from social media learning data. With appropriate extensions (more cohorts, individual-level modelling, lagged features), the approach has potential to inform real-time monitoring systems that give instructors timely, evidence-based signals about participation dynamics.

References

Anderson, L.W. and Bloom, B.S. (2001), *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Longman.

Anderson, T., Rourke, L., Garrison, D.R. and Archer, W. (2001), “Assessing teaching presence in a computer conferencing context”, *Journal of the Asynchronous Learning Network*, Vol. 5 No. 2, pp. 1-17, doi: 10.24059/olj.v5i2.1875.

Babik, D., Gehringer, E., Kidd, J., Sunday, K., Tinapple, D. and Gilbert, S. (2024), “A systematic review of educational online peer-review and assessment systems: charting the landscape”, *Educational Technology Research and Development*, Vol. 72 No. 3, pp. 1653-1689, doi: 10.1007/s11423-024-10349-x.

Berge, Z.L. (1995), “Facilitating computer conferencing: recommendations from the field”, *Educational Technology*, Vol. 35 No. 1, pp. 22-30.

Borup, J., West, R.E. and Graham, C.R. (2012), “Improving online social presence through asynchronous video”, *The Internet and Higher Education*, Vol. 15 No. 3, pp. 195-203, doi: 10.1016/j.iheduc.2011.11.001.

Bravo, C., Redondo, M.A., Verdejo, M.F. and Ortega, M. (2008), “A framework for process–solution analysis in collaborative learning environments”, *International Journal of Human-Computer Studies*, Vol. 66 No. 11, pp. 812-832, doi: 10.1016/j.ijhcs.2008.08.003.

Bruggeman, B., Tondeur, J., Struyven, K., Pynoo, B., Garone, A. and Vanslambrouck, S. (2021), “Experts speaking: crucial teacher attributes for implementing blended learning in higher education”, *The Internet and Higher Education*, Vol. 48, 100772, doi: 10.1016/j.iheduc.2020.100772.

Cercone, K. (2008), “Characteristics of adult learners with implications for online learning design”, *AACE Journal*, Vol. 16 No. 2, pp. 137-159.

Chen, F., Li, S., Lin, L. and Huang, X. (2024), “Identifying temporal changes in student engagement in social annotation during online collaborative reading”, *Education and Information Technologies*, Vol. 29 No. 13, pp. 16101-16124, doi: 10.1007/s10639-024-12494-5.

Creswell, J.W. and Creswell, J.D. (2017), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage Publications.

Delahunty, J., Verenikina, I. and Jones, P. (2014), “Socio-emotional connections: identity, belonging and learning in online interactions. A literature review”, *Technology, Pedagogy and Education*, Vol. 23 No. 2, pp. 243-265, doi: 10.1080/1475939x.2013.813405.

Dokhanchi, M., Kavanagh, L. and Reidsema, C. (2018), “Factors that influence peer learning in social media enhanced engineering courses”, *Proceedings of the 29th Annual Conference of the Australasian Association for Engineering Education*, Hamilton, New Zealand.

Fiock, H. (2020), “Designing a community of inquiry in online courses”, *The International Review of Research in Open and Distributed Learning*, Vol. 21 No. 1, pp. 135-153, doi: 10.19173/irrodl.v20i5.3985.

Freeman, M.T.M. and Jarvie-Eggart, M.E. (2019), “Best practices in promoting faculty-student interaction in online STEM courses”, *126th Annual Conference and Exposition*, Tampa, Florida.

Garrison, R. (2022), “Shared metacognition in a community of inquiry”, *Online Learning*, Vol. 26 No. 1, pp. 6-18, doi: 10.24059/olj.v26i1.3023.

Garrison, D.R. and Cleveland-Innes, M. (2005), “Facilitating cognitive presence in online learning: interaction is not enough”, *The American Journal of Distance Education*, Vol. 19 No. 3, pp. 133-148, doi: 10.1207/s15389286ajde1903_2.

Garrison, D.R., Anderson, T. and Archer, W. (1999), “Critical inquiry in a text-based environment: computer conferencing in higher education”, *The Internet and Higher Education*, Vol. 2 Nos 2-3, pp. 87-105, doi: 10.1016/s1096-7516(00)00016-6.

Hare, A.P., Blumberg, H.H., Davies, M.F. and Kent, M.V. (1994), *Small Group Research: A Handbook*, Ablex Publishing.

Hillman, D.C., Willis, D.J. and Gunawardena, C.N. (1994), “Learner–interface interaction in distance education: an extension of contemporary models and strategies for practitioners”, *American Journal of Distance Education*, Vol. 8 No. 2, pp. 30-42, doi: 10.1080/08923649409526853.

Kanuka, H. and Garrison, D.R. (2004), “Cognitive presence in online learning”, *Journal of Computing in Higher Education*, Vol. 15 No. 2, pp. 21-39, doi: 10.1007/bf02940928.

Ke, F. and Xie, K. (2009), “Toward deep learning for adult students in online courses”, *The Internet and Higher Education*, Vol. 12 Nos 3-4, pp. 136-145, doi: 10.1016/j.iheduc.2009.08.001.

Lai, C.H., Lin, H.W., Lin, R.M. and Tho, P.D. (2019), “Effect of peer interaction among online learning community on learning engagement and achievement”, *International Journal of Distance Education Technologies (IJDET)*, Vol. 17 No. 1, pp. 66-77, doi: 10.4018/ijdet.2019010105.

Lee, L. (2002), “Enhancing learners’ communication skills through synchronous electronic interaction and task-based instruction”, *Foreign Language Annals*, Vol. 35 No. 1, pp. 16-24, doi: 10.1111/j.1944-9720.2002.tb01829.x.

Lin, C.H., Zheng, B. and Zhang, Y. (2017), “Interactions and learning outcomes in online language courses”, *British Journal of Educational Technology*, Vol. 48 No. 3, pp. 730-748, doi: 10.1111/bjet.12457.

Liu, J.C. and Kaye, E.R. (2016), “Preparing online learning readiness with learner-content interaction: design for scaffolding self-regulated learning”, in *Handbook of Research on Strategic Management of Interaction, Presence, and Participation in Online Courses*, IGI Global, pp. 216-243.

Liu, W., Xie, H. and Johnson, C. (2017), “The influencing factors and effects on comprehension of e-reading”, *2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chengdu, China.

Majeski, R. and Stover, M. (2007), “Theoretically based pedagogical strategies leading to deep learning in asynchronous online gerontology courses”, *Educational Gerontology*, Vol. 33 No. 3, pp. 171-185, doi: 10.1080/03601270600850826.

Moon, J., McNeill, L., Edmonds, C.T., Banihashem, S.K. and Noroozi, O. (2024), “Using learning analytics to explore peer learning patterns in asynchronous gamified environments”, *International Journal of Educational Technology in Higher Education*, Vol. 21 No. 1, p. 45, doi: 10.1186/s41239-024-00476-z.

Moore, M. (1989), “Three types of interaction”, *The American Journal of Distance Education*, Vol. 3 No. 2, pp. 1-7, doi: 10.1080/08923648909526659.

Moore, M. (1993), “Three types of interaction”, in *Distance Education: New Perspectives*, pp. 19-24.

Ng, B.J.M., Han, J.Y., Kim, Y., Togo, K.A., Chew, J.Y., Lam, Y. and Fung, F.M. (2021), “Supporting social and learning presence in the revised community of inquiry framework for hybrid learning”, *Journal of Chemical Education*, Vol. 99 No. 2, pp. 708-714, doi: 10.1021/acs.jchemed.1c00842.

Palloff, R.M. and Pratt, K. (1999), *Building Learning Communities in Cyberspace*, Jossey-Bass, San Francisco.

Redmond, P. (2014), “Reflection as an indicator of cognitive presence”, *E-Learning and Digital Media*, Vol. 11 No. 1, pp. 46-58, doi: 10.2304/elea.2014.11.1.46.

Rourke, L., Anderson, T. and Garrison, D.R. (1999), “Assessing social presence in asynchronous text-based computer conferencing”, *Journal of Distance Education*, Vol. 14 No. 2, pp. 50-71.

Rovai, A.P. (2002), “Building sense of community at a distance”, *The International Review of Research in Open and Distributed Learning*, Vol. 3 No. 1, pp. 1-16, doi: 10.19173/irrodl.v3i1.79.

Shea, P. and Bidjerano, T. (2009), “Community of inquiry as a theoretical framework to foster ‘epistemic engagement’ and ‘cognitive presence’ in online education”, *Computers and Education*, Vol. 52 No. 3, pp. 543-553, doi: 10.1016/j.compedu.2008.10.007.

Shea, P. and Bidjerano, T. (2012), “Learning presence as a moderator in the community of inquiry model”, *Computers and Education*, Vol. 59 No. 2, pp. 316-326, doi: 10.1016/j.compedu.2012.01.011.

Shute, V.J. (2008), “Focus on formative feedback”, *Review of Educational Research*, Vol. 78 No. 1, pp. 153-189, doi: 10.3102/0034654307313795.

Stenbom, S. (2018), "A systematic review of the community of inquiry survey", *The Internet and Higher Education*, Vol. 39, pp. 22-32, doi: 10.1016/j.iheduc.2018.06.001.

Swan, K. and Shih, L.F. (2005), "On the nature and development of social presence in online course discussions", *Journal of Asynchronous Learning Networks*, Vol. 9 No. 3, pp. 115-136, doi: 10.24059/olj.v9i3.1788.

Vuopala, E., Hyvönen, P. and Järvelä, S. (2016), "Interaction forms in successful collaborative learning in virtual learning environments", *Active Learning in Higher Education*, Vol. 17 No. 1, pp. 25-38, doi: 10.1177/1469787415616730.

Wagner, E.D. (1997), "Interactivity: from agents to outcomes", *New Directions for Teaching and Learning*, Vol. 1997 No. 71, pp. 19-26, doi: 10.1002/tl.7103.

Xie, V.D. and Yen, L.L. (2011), "Relationship between students' motivation and their participation in asynchronous online discussions", *MERLOT Journal of Online Learning and Teaching*, Vol. 7 No. 1, pp. 17-29.

Xie, H., Liu, W. and Bhairma, J. (2018), "Analysis of synchronous and asynchronous e-learning environments", *3rd Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC)*, Chongqing, China.

Yu, Q. and Schunn, C.D. (2023), "Understanding the what and when of peer feedback benefits for performance and transfer", *Computers in Human Behavior*, Vol. 147, 107857, doi: 10.1016/j.chb.2023.107857.

Zhang, E., Zhang, Z., Liu, H., Han, S. and Xue, Z. (2025), "Exploring peer facilitation and critical thinking in asynchronous online discussions: a lag sequential analysis approach", *British Journal of Educational Technology*, pp. 2453-2477, doi: 10.1111/bjet.13583.

Zhu, E. (2006), "Interaction and cognitive engagement: an analysis of four asynchronous online discussions", *Instructional Science*, Vol. 34 No. 6, pp. 451-480, doi: 10.1007/s11251-006-0004-0.

Author Affiliations

Mohsen Dokhanchi, Faculty of Engineering Architecture and Information Technology, The University of Queensland, Saint Lucia, Australia

Shahrzad Saremi, School of Science, Technology and Engineering, University of the Sunshine Coast, Sunshine Coast, Australia

Rania Shibl, Faculty of Science and Engineering, Southern Cross University, Gold Coast, Australia

Maryam Heidari, Griffith University School of Information and Communication Technology, Gold Coast, Australia

Hassan Ahmed, Department of Computer Science, National University of Computer and Emerging Sciences - Chiniot-Faisalabad Campus, Chiniot, Pakistan

Dahlia Mansoor, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Yassine Himeur, Abu Dhabi University College of Business Administration, Abu Dhabi, United Arab Emirates

Mohammad Al-Zaffin, Business School, University of Dubai, Dubai, United Arab Emirates and University of Dubai, Dubai, United Arab Emirates

Shadi Atalla, College of Engineering and Information Technology, University of Dubai, Dubai, United Arab Emirates

Wathiq Mansoor, American University of Iraq, Sulaimani, Iraq

Corresponding author

Shahrzad Saremi can be contacted at: ssaremi@usc.edu.au