

# Predicting Cognitive Engagement in Online Course Discussion Forums

Guher Gorgun  
Educational Psychology  
University of Alberta  
Edmonton, AB, Canada  
gorgun@ualberta.ca

Seyma N. Yildirim-Erbasli  
Educational Psychology  
University of Alberta  
Edmonton, AB, Canada  
seymanur@ualberta.ca

Carrie Demmans Epp  
EdTeKLA, Computing Science  
University of Alberta  
Edmonton, AB, Canada  
cdemmansepp@ualberta.ca

## ABSTRACT

The need to identify student cognitive engagement in online-learning settings has increased with our use of online learning approaches because engagement plays an important role in ensuring student success in these environments. Engaged students are more likely to complete online courses successfully, but this setting makes it more difficult for instructors to identify engagement. In this study, we developed predictive models for automating the identification of cognitive engagement in online discussion posts. We adapted the Interactive, Constructive, Active, and Passive (ICAP) Engagement theory [15] by merging ICAP with Bloom's taxonomy. We then applied this adaptation of ICAP to label student posts ( $N = 4,217$ ), thus capturing their level of cognitive engagement. To investigate the feasibility of automatically identifying cognitive engagement, the labelled data were used to train three machine learning classifiers (i.e., decision tree, random forest, and support vector machine). Model inputs included features extracted by applying Coh-Metrix to student posts and non-linguistic contextual features (e.g., number of replies). The support vector machine model outperformed the other classifiers. Our findings suggest it is feasible to automatically identify cognitive engagement in online learning environments. Subsequent analyses suggest that new language features (e.g., AWL use) should be included because they support the identification of cognitive engagement. Such detectors could be used to help identify students who are in need of support or help adapt teaching practices and learning materials.

## Keywords

cognitive engagement, Coh-Metrix, discussion forums, natural language processing (NLP), online learning

## 1. INTRODUCTION

G. Gorgun, S. N. Yildirim-Erbasli, and C. D. Epp. Predicting cognitive engagement in online course discussion forums. In A. Mitrovic and N. Bosch, editors, *Proceedings of the 15th International Conference on Educational Data Mining*, pages 276–289, Durham, United Kingdom, July 2022. International Educational Data Mining Society.

© 2022 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.  
<https://doi.org/10.5281/zenodo.6853149>

Educational theories and empirical studies have emphasized the importance of learner engagement [34]. Engaged learners are active; they invest time and effort during the learning process [44, 56, 66]. Empirical studies have corroborated the importance of engagement for student learning and well-being [39, 57, 74]. Studies have consistently found a significant association between student engagement and the cognitive and non-cognitive skills of students. For example, engaged students are more likely to be academically successful [89], retain knowledge [9], and show higher levels of critical-thinking [13] and problem-solving [61, 97]. Furthermore, engaged learners are more likely to show higher levels of school belongingness and socio-emotional well-being [22, 36, 73]. They are also less likely to drop-out [2, 3, 41] or demonstrate negative behaviors in school [62, 93]. Consistent with these general educational outcomes, a positive association between active participation in online discussion forums and academic achievement has been found [23, 51, 85]. Learner engagement, therefore, is a vital and predictive aspect of student learning and well-being [30].

Given that student engagement is a strong correlate of student achievement, researchers have long been interested in identifying student engagement levels to improve student success and learning processes. Identifying student engagement levels could be used to generate actionable feedback for students [52]. This feedback can be provided by instructors or through automated feedback that is delivered via learning dashboards or system features that nudge students to engage with the task.

However, identifying student (dis)engagement is difficult for instructors in online settings. In recent years, researchers have been interested in automating engagement identification. Automatically identifying students' engagement levels may also help instructors identify potentially at-risk students. Instructors may intervene to minimize the negative impact of disengagement on student learning. Furthermore, engagement detection could generate real-time feedback that instructors could use themselves. This critical feedback may be used to inform and evaluate the effectiveness of instructional practices employed in the class. Additionally, engagement detection could be embedded in instructor-facing analytics dashboards to help them understand student engagement in their classes [67]. Detecting that many learners are disengaged with the task could help instructors change their instructional strategy and practice, adjusting to meet the

needs of learners in their class [1, 68]. Providing instructors with information about learner engagement could also enable them to intervene in a timely manner [31].

Instructors can harness information that is easily obtained in face-to-face learning environments to observe student behavior and make inferences about how they are engaging. However, certain types of engagement (e.g., cognitive, affective) are not directly observable. Students may also feign engagement. Even if the form of engagement the instructor is interested in monitoring is easy to observe in person, that may not be the case in online settings. With the recent challenges that the COVID-19 pandemic has brought and the associated widespread use of online learning environments [69], the automated detection of learner engagement has become more vital yet more complicated. With the use of log data and dashboards, instructors could evaluate learner time within the system or time on task as an indication of behavioral engagement [67]. Unfortunately, proxies, such as time spent within a system, will not help an instructor extract the type of cognitive engagement that students demonstrate in online learning environments.

Many have tried to define student engagement, which is latent and inherently multi-faceted. Along with this, they have tried to model different types of learner engagement (e.g., academic, behavioral, affective, cognitive) in various learning contexts (e.g., online education, in-person classes). Typically, engagement is defined as students' motivation, willingness, effort, and involvement in school or learning-related tasks [19, 87]. It is operationalized as students' investment of resources such as effort, time, and energy into learning [53, 60, 61, 62].

Behavioral, emotional, and cognitive engagement are frequently studied types of engagement. Behavioral engagement is generally defined as time on task [18, 42, 50, 64, 94], and focuses on overt student behaviors [7]. Emotional engagement typically refers to learners' feelings, affect, and emotional reactions [77, 81, 83]. Cognitive engagement is conceptualized as employing learning strategies and investing effort and persistence into the task at hand [18, 42, 86, 94]. Different types of engagement are not mutually exclusive. For example, some researchers have argued that cognitively engaged learners are also behaviorally engaged, but the reverse is not necessarily true [50].

Cognitive engagement is one of the most challenging types of engagement to detect. Although there are certain theories developed for characterizing and capturing cognitive engagement in physical classroom settings (e.g., ICAP; [15]), they can be difficult to directly apply in online discussion environments, particularly in discussion forum posts (see [90]).

In this study, we focused on developing predictive models for automating cognitive engagement detection in discussion posts using natural language processing (NLP) and machine learning (ML) methods. This will help us evaluate the feasibility of automating cognitive engagement identification in online discussion posts. We will also analyze the classification errors and feature importance when predicting cognitive engagement through these relatively transparent models. Understanding feature importance and classifica-

tion errors may help us develop better models for detecting cognitive engagement.

## 2. ENGAGEMENT IDENTIFICATION APPROACHES

Dewan and colleagues [30] proposed an engagement detection taxonomy and grouped engagement detection methods under three main categories: manual (e.g., self-report measures), semi-automatic (e.g., engagement tracing), and automatic (e.g., log file analysis).

Manual engagement detection methods include self-report measures, checklists, and rating scales. Using self-report measures, students are asked to indicate their own engagement levels (e.g., [5, 12, 43]). Although it is easier to administer and use self-report measures [82], one major limitation of these instruments is their susceptibility to reporter biases. When learners rate their own engagement level, they may not accurately reflect their engagement during learning processes [35] because they lack the capacity to do so, or students may not reflect their true engagement status to avoid getting in trouble in a class (e.g., [45]). Therefore, self-report measures may not provide an objective evaluation of engagement.

The checklists and rating scales that instructors use are less prone to self-report biases (e.g., self-serving bias) during the coding process. Yet, checklists and rating scales also have certain limitations. Using these methods requires a great deal of time and effort [30]. For example, instructors, especially in large-scale classrooms or online environments, may not track each learner accurately. Learners may also pretend to be on task when in fact they are not engaged with the task. Hence, teacher evaluations made through such means may not accurately present the engagement levels of learners in class. Additionally, both self-report measures and checklists may not accurately capture fluctuations in learner engagement [90].

Semi-automatic engagement detection methods include engagement tracing. Using log or trace data, the timing and accuracy of learner responses have been employed to identify engaged and disengaged learners [8, 54]. For example, an unrealistically short response time is considered an indication of student disengagement during a quiz attempt.

Finally, automatic engagement detection methods include sensor data, log data, and computer vision-based approaches that analyze eye movement, body posture, or facial expressions [30]. These detection methods may automatically extract features from learners' body movement or facial expressions using sensor data as input [91]. Additionally, learner activities have been traced in learning management systems to automatically extract features related to engagement detection [16, 17]. These approaches may provide useful real-time information about student engagement, but they introduce privacy concerns such as being recorded.

### 2.1 Cognitive Engagement Frameworks

Two frequently employed cognitive engagement frameworks include Community of Inquiry (CoI; [46]) and Interactive, Cognitive, Active, and Passive engagement (ICAP; [15]).

The CoI framework identifies three elements (i.e., social presence, teaching presence, and cognitive presence) that support successful learning. The cognitive presence element of the CoI framework has been widely used by researchers (e.g., [58, 70]) for analyzing student learning and developing predictive models in online courses. In those studies, cognitive presence encompasses five phases: triggering event, exploration, integration, resolution, and other.

The ICAP framework [15] conceptualizes hierarchical cognitive engagement levels. Higher levels are related to higher cognitive engagement and learning growth. From top to bottom, the order of cognitive engagement levels is interactive, constructive, active, and passive.

Although both frameworks have been used to model engagement, they approach the categorization of engagement differently and have distinct aims [38]. The CoI framework specifically targets online learning environments and tries to model how students develop ideas in online discussions. The ICAP framework has been used to characterize learning in both in-person and online learning environments. It focuses on students' *active learning* behaviors [15]. Both frameworks appear to be promising theoretical approaches for cognitive engagement modeling and prediction. CoI focuses on different phases whereas ICAP focuses on different levels of cognitive engagement.

In a recent study, Farrow and colleagues [38] compared these engagement frameworks to decipher their commonalities and differences. They found similarities between the predictors used for engagement detection (e.g., message length was correlated with higher levels of engagement in both frameworks) whereas there were differences in the interpretation of classes (e.g., ICAP rewarded interactivity more than CoI).

## 2.2 Identifying Engagement in Online Environments

Some studies of online learning have used qualitative content analysis methods [90, 98] to detect engaged and disengaged learners. These studies start with developing a coding scheme reflecting different levels of engagement (e.g., active or passive engagement). This coding scheme is then used by the trained coders to manually label learner activities. This labelling process requires that considerable time and manual labor be invested, which is why it is not surprising that automating engagement identification via machine learning methods has emerged as a viable approach.

There are several potential benefits of automating cognitive engagement detection that go beyond the reduction in human fallibility and manual labor. These potential benefits include reducing coding time, rapidly identifying at-risk students in terms of engagement, and the possibility of integrating learner engagement information into dashboards and learning management systems. Moreover, automation can build on the prior work that manually labelled student engagement and helped us to understand it at a smaller scale.

Automated engagement detection methods in online environments can be divided into two groups [56]. The first group of studies focuses on automatically extracting features

pertaining to learners' physical cues such as body movement, heart rate, head posture, or where learners are looking (e.g., [55, 65]). In a recent study, Li and colleagues [65] extracted students' facial features and trained a supervised model to classify student cognitive engagement type. Although facial features were found to be powerful predictors of cognitive engagement, these types of studies may require webcams to extract learners' facial features, body posture, or head postures. Additionally, the learners could be aware of the fact that they are being recorded, which can lead them to experience discomfort [29]. Furthermore, the implementation of such systems may be costly (e.g., integrating webcams) compared with extracting learner trace data from learning management systems.

The second group of studies focuses on using a less sensor-heavy approach. They extract linguistic features (e.g., coherence, number of words) and trace data (e.g., time on task, number of clicks) from online learning environments to detect engagement (e.g., [6, 58, 59, 70]). For example, Kovanović and colleagues [58] employed n-grams and part-of-speech features to train a predictive model of cognitive presence using the CoI framework [46]. Their model achieved 58.38% accuracy ( $K = .41$ ). In a similar study, Kovanović and colleagues [59] extracted features using Coh-Metrix, Linguistic Inquiry and Word Count (LIWC), and latent semantic analysis (LSA) similarity to represent average sentence similarity. Their best model achieved 72% accuracy ( $K = .65$ ). Moving from the COI lens to that of ICAP, Atapattu and colleagues [6] used word embeddings from Doc2Vec to detect only the active and constructive levels of cognitive engagement in a MOOC, where they ignored posts that were of a social nature. They argued that learners with active engagement posts paraphrased, repeated, or mapped resources whereas learners with constructive posts proposed new ideas or introduced external material going beyond what was covered in class [6]. Therefore, discussion posts similar to course content remained in close proximity to the vector space generated by Doc2Vec. These posts were classified as active engagement whereas discussion posts that were far away from the vector space were classified as constructive engagement.

Some linguistic and course-based contextual features might limit the generalizability of predictive models (e.g., n-gram based models are sensitive to vocabulary choices) of cognitive engagement across different courses and contexts. For example, in a recent study, Neto and colleagues [70] analyzed model generalizability across educational contexts by employing Coh-Metrix features and the CoI framework. In their study, the baseline model was trained with a biology course dataset and achieved 76% accuracy ( $K = .55$ ). When applied to a dataset from a technology course, the model had an accuracy of 67% ( $K = .20$ ), which indicates some limitations to the generalizability of the model to new courses.

## 3. PRESENT STUDY

The purpose of this study is to evaluate the feasibility of automating cognitive engagement identification in fully online graduate courses through the lens of the ICAP framework. Given that engagement is a latent trait, researchers typically start with creating labeled data so that they can train supervised ML models. Most studies have employed CoI for

labeling cognitive presence (e.g., [37, 58, 70]).

ICAP and CoI have different aims and operationalizations of cognitive engagement (see section 2.1). In this study, we defined cognitive engagement as investing effort and cognitive resources. ICAP is more congruent with our cognitive engagement definition since we want to identify different levels of engagement in posts. Therefore, we employed the ICAP framework as our theoretical background that informed labeling decisions.

There is a lack of consensus concerning the coding scheme of ICAP for online environments even though the available coding schemes partly overlap with one another. For example, while Atapattu et al. [6] employed a binary coding scheme (i.e., active vs constructive), Yogeve et al. [95] employed a coding scheme with six categories.

We first adapted the available coding schemes for online discussion posts by aligning the engagement levels with Bloom’s taxonomy [4]. This was done to emphasize the increased nature of cognitive complexity in both the taxonomy and cognitive engagement. Then, we developed three supervised ML models to analyze the feasibility of cognitive engagement prediction in online discussion environments. We also analyzed feature importance to evaluate whether we identified the same order of features across the classifiers employed in this study.

Our research questions were as follows:

**RQ1:** To what extent can a model trained with Coh-Metrix and contextual features be used to automatically classify discussion posts based on cognitive engagement level?

**RQ2:** Which features are more important for cognitive engagement prediction?

**RQ3:** What types of misclassifications occur?

## 4. METHODS

The data for this study consists of discussion forums from fully-online, graduate-level courses. The course forums differed in terms of facilitation method (i.e., peer-facilitated vs instructor-facilitated) and course length, i.e., regular-term courses (long) and summer-term courses (short).

### 4.1 Participants and Study Context

The data used in this study were collected from an online discussion platform that is used to deliver courses in a highly ranked college of education in Canada. The data collection protocol was reviewed and approved by the university’s research ethics board prior to completion. Participant consent was obtained for supplementary data collection. The dataset included 4,217 posts that had been produced by 111 students. In Table 1, we present the number of students, term length, facilitation method, and percentage of cognitive engagement posts in each course.

The courses span departments within the college and cover topics from language learning to educational psychology, educational technology, and educational policy.

**Table 1: Number of students, term length, facilitation method, and percentage (%) of cognitive engagement levels by course.**

ID	<i>n</i>	Length	Method	S	A	C	I
1	31	Long	Instructor	57	24	17	2
2	18	Short	Instructor	10	30	41	19
3	17	Short	Peer	44	22	12	22
4	22	Long	Instructor	24	31	23	24
5	23	Long	Peer	11	29	17	43

Note. Social (S), Active (A), Constructive (C), Interactive (I)

Demographic data were not collected and the previous deidentification of the data means that it cannot be obtained retroactively. It is worth noting that students could have their data excluded at the post level by marking it as private.

### 4.2 Data Coding Procedure

Our coding scheme is largely informed by the coding framework developed by Wang et al. [90] and Yogeve et al. [95]. We altered and simplified the coding scheme based on the challenges we observed during the coding process.

We informed the coding scheme with Bloom’s revised taxonomy [4] and adapted it to reflect the higher-order cognitive complexity in the ICAP framework. Specifically, we map Bloom’s taxonomy level indicators onto cognitive engagement indicators. Bloom’s taxonomy was chosen because it is widely used by K-12 and higher education instructors to develop measurable and observable instructional objectives, tapping different levels of cognitive complexity. Cognitive complexity refers to the amount of cognitive demand required to complete a task. Higher levels of cognitive engagement necessitate higher effort investment and align with higher levels of cognitive complexity in Bloom’s taxonomy. Hence, we emphasized the presence of higher-order skills when identifying higher levels of cognitive engagement in the coding scheme.

Figure 1 shows the cognitive engagement coding scheme we used for labelling the discussion forum posts.<sup>1</sup> It highlights how the hierarchical nature of cognitive engagement aligns with Bloom’s taxonomy levels. For example, in the lower levels of the taxonomy, we have *remembering* and *understanding*. Active engagement corresponds to these levels in the taxonomy since posts with active engagement include elements of paraphrasing, mapping resources, and retrieving the same or similar concepts that are covered in class. Constructive engagement aligns with the level of *applying* and *analyzing* in Bloom’s taxonomy since these posts compare, contrast, illustrate, and argue in a cause and effect fashion. Finally, Interactive engagement relates to the levels of *evaluating* and *creating* in the taxonomy because these discussion posts make judgments and evaluations about the topics covered.

Our coding scheme included four categories: social, active, constructive, and interactive. The engagement categories are hierarchical, and social engagement is the lowest cate-

<sup>1</sup>We cannot share the post content due to learner privacy.

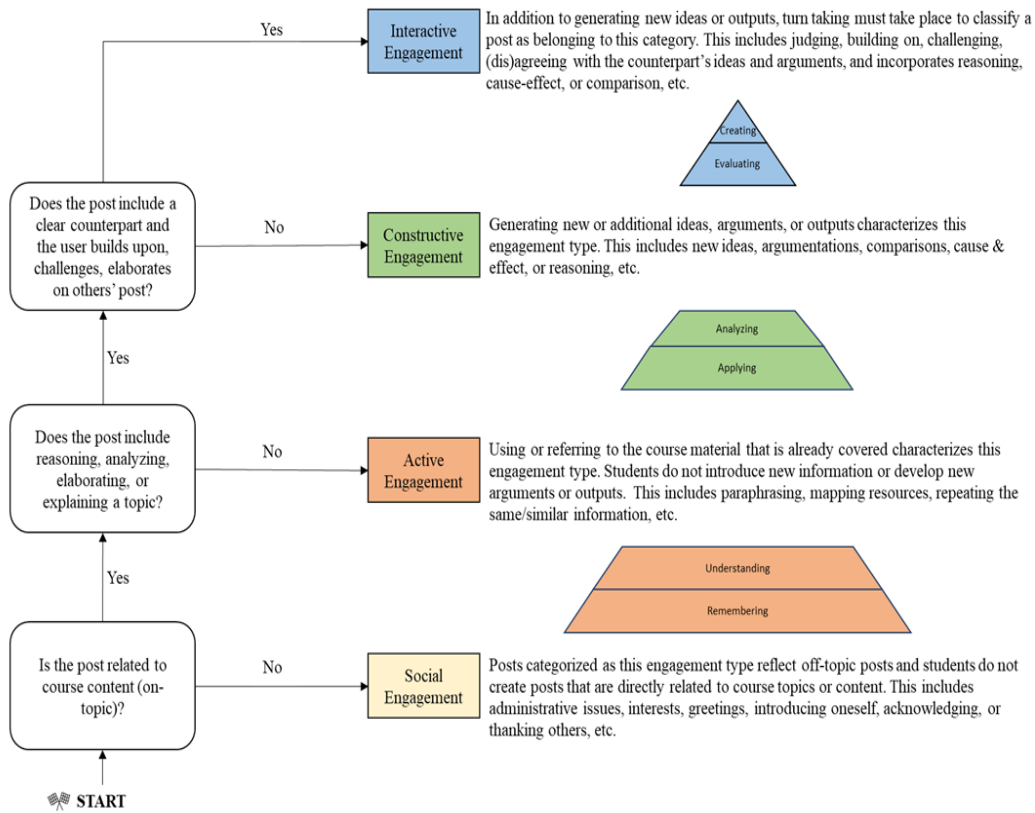


Figure 1: Cognitive engagement coding scheme.

gory followed by active, constructive, and interactive engagement. Note that we did not use *passive* engagement since our goal is not to categorize students' engagement levels but rather to categorize the engagement observed in posts. Passive engagement entails that students read the posts but do not produce any posts, hence it cannot be employed for categorizing engagement in discussion posts. We also introduced a *social* engagement level corresponding to posts that are off-task and are meant to support relationship building (e.g., students introducing themselves).

Based on the coding scheme, we first identified whether the posts were on-task or off-task. Then, we asked whether the post exhibited reasoning, argumentation, or elaboration on a topic. Finally, we asked whether there was an evaluative argument and a clear counterpart in each post. While coding the posts, we assigned the highest engagement level observed in each post.

The coding was done by two content experts (Authors 1 and 2) who had previous experience with data coding and who were familiar with the cognitive engagement literature. First, we had a training session where we coded a sample of discussion posts, identified challenges, and reconciled discrepancies during the coding process. After the training session, we each coded a sample of discussion posts independently and we checked agreement between raters using the percent agreement method. We had an inter-rater reliability of .91.

Table 2: Distribution of cognitive engagement levels

ID	Label	Engagement Type	Number (%)
0		Social	1197 (28)
1		Active	1173 (28)
2		Constructive	930 (22)
3		Interactive	917 (22)

Table 2 shows the distribution of the four levels of cognitive engagement across all posts. The most frequently observed levels of cognitive engagement were social (28%) and active (28%). Given that constructive and interactive levels of engagement require higher cognitive effort and complexity, we expected to observe fewer constructive and interactive engagement level posts. Even though we expected to observe class imbalances based on previous studies (e.g., [38, 70]), we had a nearly balanced distribution of cognitive engagement levels. This might be due to the class structure where some instructors relied on peer facilitation as a management strategy for the course forums. Alternatively, having graduate-level courses could explain this balance.

### 4.3 Feature Extraction

This study used Coh-Metrix and non-linguistic contextual features to classify the cognitive engagement level seen in discussion posts.

Coh-Metrix is a tool used for discourse analysis. It estimates the cohesion, coherence, linguistic complexity, read-

**Table 3: Hyperparameter search space used for model tuning**

ML Classifier	Parameters	Values
Decision Tree	Max depth	3, 6, 9, 12, None
	Min samples leaf	A random integer between 1 and 10
	Max features	A random integer between 1 and 10
	Criterion	'Gini', 'Entropy'
Random Forest	Max depth	3, 6, 9, 12
	Min samples leaf	2, 4, 6, 10
	Max features	5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 107
	Number of estimators	500, 1000
	Min samples split	5, 10
	Criterion	'Gini', 'Entropy'
	Bootstrap	'True', 'False'
Support Vector Machine	Kernel	'linear', 'sigmoid', 'poly', 'rbf'
	C	50, 10, 1, 0.1

ability, and lexical category use in a text [32]. The English version of Coh-Matrix incorporates 108 features including referential cohesion, deep cohesion, and narrativity [33]. In previous studies on text classification, Coh-Matrix and contextual features were shown to be promising for cognitive presence identification (i.e., Community of Inquiry; [70]), suggesting their potential for supporting cognitive engagement detection in online discussion environments.

In addition to the Coh-Matrix features, we included three contextual features. These contextual features included information about whether a discussion post is a reply to another post, the number of replies that the post had received, and a count of the use of vocabulary from the academic word list (AWL) [20]. Whether a post is a reply was a binary variable, the number of replies a post received was an integer, and AWL count was an integer representing the total number of academic words in a post.

#### 4.4 Data Pre-processing

To clean and prepare the corpus for training and classification, we ran several pre-processing steps. Our goal was to build a cognitive engagement classifier, thus we tried to create a corpus that was as clean and close to human readable form as possible [32]. We removed website links and “see attached” notations; stripped the html tags, eliminated new lines, white spaces, and tabs; expanded contractions; removed numbering and bullet points; and corrected misspelled words.

Because our unit of analysis was discussion posts, we created separate files for each post after applying the above data cleaning steps. We then ran Coh-Matrix 3.0 [47] to extract linguistic features for each post. All discussion posts have a single paragraph, so we removed the Coh-Matrix indicators of paragraph count (i.e., DESPC), standard deviation of the mean length of paragraphs (i.e., DESPLd), the mean of the LSA cosines between adjacent paragraphs (i.e., LSAPP1), and the standard deviation of LSA cosines between adjacent paragraphs (i.e., LSAPP1d). Finally, we created a data file containing all of the input features and the class labels. The input features included the 104 Coh-Matrix indicators and three other non-linguistic contextual features (e.g., whether a post is a reply). This data file was then used to train our

classifiers for cognitive engagement prediction.

#### 4.5 Model Selection

To train and test models for predicting cognitive engagement, we split the dataset in two: 70% was used as a training set and the remaining 30% was used as a test set.

To answer our research questions, we trained three types of supervised classifiers:

- a **decision tree** (DT) [10] was selected because this approach is easier to interpret and the graphical representation of the tree can help us understand the relative importance of features;
- a **random forest classifier** (RF) [80] was chosen because it is an ensemble method that often exhibits superior performance on classification tasks in educational contexts (e.g., [59, 70]); and
- the **support vector machine** (SVM) [21] algorithm was chosen because it is designed to handle multidimensional data which may lead to superior performance when predicting cognitive engagement in discussion posts [40].

In addition to the above attributes, these types of models have previously performed well in other forum classification tasks [88] or they have supported educational data mining tasks with similarly sized or smaller data sets [25, 78]. Moreover, these classification algorithms are relatively transparent so they can aid us in understanding the contribution of each feature to cognitive engagement prediction.

#### 4.6 Model Tuning

Model training, hyperparameter tuning, and analyses were conducted in Python (Version 3.8.8) using the sklearn [71] and mlxtend [75] packages.

We tuned the hyperparameters of each classifier using randomized search with 10-fold nested cross-validation. In Table 3, we summarize the hyperparameters and values used to tune the models for each classifier. For the decision tree,

**Table 4: Model performance by model and cognitive engagement level (Cog. Engage.)**

Cog. Engage.	Decision Tree				Random Forest				Support Vector Machine			
	P	R	F1	Acc. (%)	P	R	F1	Acc. (%)	P	R	F1	Acc. (%)
Social	.65	.70	.68	71	.77	.74	.76	73	.79	.83	.81	83
Active	.50	.46	.48	46	.56	.56	.56	52	.60	.54	.57	54
Constructive	.71	.58	.64	62	.74	.69	.71	68	.77	.73	.75	73
Interactive	.55	.65	.59	64	.60	.68	.64	70	.64	.73	.68	73
Full	.59	.58	.58	60	.66	.66	.65	66	.70	.71	.70	71

the best model performance was obtained with max depth = 6, criterion = Gini, max features = 106, and min samples leaf = 7. For RF, the best model performance was obtained with max depth = 40, criterion = entropy, and number of estimators = 600. The best performing SVM model was obtained with kernel = linear and regularization (C) = 0.1.

The best model achieved 68% accuracy for the decision tree, 95% accuracy for random forest, and 85% for SVM on the training set.

#### 4.7 Model Comparison and Analysis

We evaluated classifier performance with the test set using accuracy (Acc), Cohen’s Kappa (K), precision (P), recall (R), and F1 using 10-fold cross-validation. We use Landis and Koch’s guidelines to interpret Cohen’s Kappa, where values below .20 indicate slight agreement, values between .21 and .40 indicate fair agreement, values between .41 and .60 indicate moderate agreement, values between .61 and .80 indicate substantial agreement, and values greater than .81 indicate strong agreement [63].

We also statistically compared the model performance using Cochran’s  $Q$  test. Cochran’s  $Q$  test is a generalized version of McNemar’s test and it can be used to compare more than two classifiers [76]. The null hypothesis for Cochran’s  $Q$  test states that there is no difference between model classification accuracies. We also used McNemar’s test, a non-parametric statistical test, to perform the subsequent paired comparisons [76]. We report continuity corrected  $p$ -values for paired comparisons.

We analyzed feature importance across models to evaluate feature contribution towards cognitive engagement identification. For the decision tree, we evaluated feature importance with the Gini index. For random forest, we used the Mean Gini Decrease value to evaluate the features with the most explanatory power. For the support vector machine, after hyperparameter tuning with kernel and regularization parameter (C), we evaluated the feature importance by comparing the size of the support vector coefficients with one another.

#### 4.8 Error Analysis Procedures

To better understand the classification errors, we analyzed the confusion matrix and the misclassified discussion posts. The confusion matrix allowed us to analyze the number of errors across models. We also compared the error rates (instead of the absolute number of errors) by dividing each value in the confusion matrix by the number of posts in the corresponding class and depicted the model performance in Figure 3. The rows represent the actual (human-assigned)

labels; the columns represent the predicted labels. This procedure allowed us to, first, identify which types of cognitive engagement tend to be misclassified as another type. Second, we compared the descriptive statistics of predictors of those misclassified posts to examine the reasons for misclassification.

## 5. RESULTS

### 5.1 RQ1: Model Performance

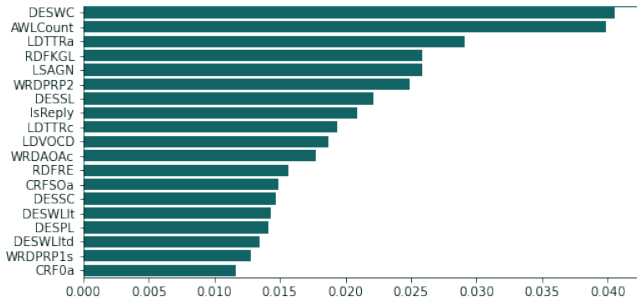
In Table 4, we provide the performance measures (i.e., precision, recall, F1 score, and accuracy) for all three models by cognitive engagement level. We also provide these measures for the full model (i.e., when all classes are being considered). Note that the accuracy of the zero-rule classifier for the full model would be 28.4%. As can be seen in Table 4, all of the models outperformed this simple baseline. Moreover, Cochran’s test revealed statistically significant differences between the classifiers we built,  $Q = 55.68$ ,  $p < .001$ . We report the results of specific paired model comparisons below.

For the decision tree, we obtained 60% accuracy (K = .46) for the full model. If we consider the F1 scores for each level of cognitive engagement within the decision tree model, we can see that it did a relatively good job of predicting social and constructive engagement. Whereas, active engagement scored below .5 on all of precision, recall, and F1. While this is better than chance, these are the lowest performance measures observed across all three models.

The accuracy of the random forest classifier was higher than that of the decision tree (McNemar’s  $\chi^2 = 21.70$ ,  $p < .001$ ). However, this increased accuracy was not accompanied by a change in agreement-level (K was .55). Again, we consider the model’s relative performance across levels of cognitive engagement, which showed that it performed best when predicting social and constructive engagement. It also performed relatively well for interactive engagement.

The support vector machine classifier outperformed both the decision tree (McNemar’s  $\chi^2 = 61.31$ ,  $p < .001$ ) and the random forest (McNemar’s  $\chi^2 = 19.04$ ,  $p < .001$ ) models on the full prediction task. The SVM model’s Kappa value (.61) suggested substantial agreement between the predicted and human-assigned labels [63]. Similar to the models trained with the random forest and decision tree algorithms, we obtained the best model performance for social and constructive engagement followed by interactive and active engagement.

### 5.2 RQ2: Feature Importance



**Figure 2: The feature importance scores for the 20 most important features for cognitive engagement prediction using random forest.**

To evaluate feature importance for predicting cognitive engagement and answer our second research question, we first analyzed the decision tree classifier. The classification tree suggested that the most important predictor was the academic word list count (AWL Count), followed by whether a post is a reply, Flesch-Kincaid grade level (i.e., Coh-Metrix indicator of RDFKGL), and number of words (i.e., DESWC from Coh-Metrix).

To interpret the random forest classifier, we used the mean decrease in Gini coefficient because it supports the evaluation of the contribution of each feature to the model’s prediction [48]. Higher values of mean decrease in Gini score indicate that the feature contributed more when performing the prediction. Similar to the decision tree model, we found that the most important features were the number of words (i.e., DESWC), number of academic word list items (AWLCount), type-token ratio for all words (LDTTRa), and Flesch-Kincaid grade level (RDFKGL). Figure 2 shows this model’s most important twenty features and their relative weights.

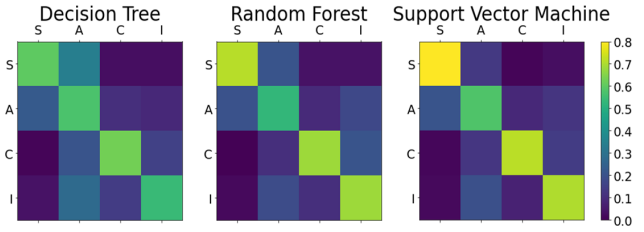
For the support vector machine model, we evaluated the coefficient importance and found that the most important feature was second language readability (RDL2). This feature was followed by type-token ratio (LDTTRc), the standard deviation of the number of syllables in words (DESWLsyd), and LSA overlap between verbs (SMCAUSlsa).

We found that the decision tree and random forest models identified the same features as most important, yet their relative contribution to the prediction task changed across models. The support vector machine, on the other hand, identified a different set of features as most important.

### 5.3 RQ3: Error Analysis

All classifiers provided better prediction performance for social and constructive engagement (see Table 4). Table 5 presents the confusion matrix for the classifiers. The superior performance of the support vector machine model is evident in Table 5, where the misclassification of posts was the lowest across engagement categories.

We analyzed the performance measures and evaluated precision, recall, and F1 scores for each classifier and engagement category. This showed the jump between the precision and



**Figure 3: Classification error rates for decision tree, random forest, and support vector machine.**

recall measures for constructive and interactive levels for the decision tree. For decision tree, the true positive rate was 65% for the interactive posts; that is, the classifier missed 35% of the interactive posts that should have been labelled with interactive. On the other hand, 45% of the posts were incorrectly identified as interactive. Furthermore, the true positive rate was 71% for the constructive posts. That is, the decision tree missed 29% of the constructive posts that should have been labelled with constructive, and 58% of the posts with other engagement levels were incorrectly identified as constructive.

The length of the posts and the smaller AWL count seem to be the primary reasons why social and active engagement posts were confused with each other. For misclassified social engagement posts, the mean post length ( $M_S = 134$ ), AWL count ( $M_S = 19.25$ ), and Flesch-Kincaid grade level ( $M_S = 12.4$ ) were similar to the mean post length ( $M_A = 138$ ), AWL count ( $M_A = 12.93$ ), and Flesch-Kincaid grade level ( $M_A = 11$ ) of active engagement posts. For the misclassified interactive posts, we observed a similar pattern. The mean post length ( $M_I = 129.5$ ), AWL count ( $M_I = 17.9$ ), and Flesch-Kincaid grade level ( $M_I = 12.65$ ) were similar to those extracted from the active posts.

Additionally, the confusion matrix (Table 5) suggested that all classifiers tend to misclassify social engagement with active engagement. Of those misclassified posts, approximately 70% of social posts were classified as active engagement across the predictive models. That is, social engagement was incorrectly grouped into the higher level. The active engagement level tended to be misclassified as social engagement, indicating model confusion between adjacent engagement categories. Of the misclassified active posts, approximately 55% of them were classified as social engagement across all predictive models. In contrast, constructive engagement tended to be misclassified as interactive engagement across all classifiers. Of the misclassified constructive posts, approximately 40% of them were classified as interactive engagement across all predictive models. Effectively, this means that it was incorrectly grouped into the higher level. Misclassifications of interactive engagement went in the opposite direction and were not for the neighbouring class. Rather, these errors were classified as active engagement instead of their being recognized as interactive posts. Of the misclassified interactive posts, approximately 60% of them were classified as active engagement across all predictive models. One possible explanation for this misclassification is as follows: Active and interactive discussion posts tended to be shorter in length compared with those at the



**Table 5: Confusion Matrix for Decision Tree, Random Forest, and Support Vector Machine**

		Decision Tree				Random Forest				Support Vector Machine				
		Actual												
Predicted		S	A	C	I	S	A	C	I	S	A	C	I	
		Social (S)	244	72	11	22	258	66	12	13	289	44	6	10
		Active (A)	89	158	33	62	69	190	30	53	67	185	36	54
		Constructive (C)	22	38	175	65	1	33	206	60	5	30	218	47
		Interactive (I)	18	49	29	179	6	53	29	187	4	47	23	201

constructive engagement level. Although constructive posts, in general, had higher word counts ( $M_C = 339$ ) and AWL counts ( $M_C = 40.3$ ), some of the misclassified active posts were shorter. This conflicted with the fact that constructive posts expanded course content with reflections and argumentation and were thus expected to be longer with greater use of academic vocabulary. This expectation did not hold in all cases and may explain why constructive engagement posts were misclassified as interactive.

To further evaluate misclassification errors for each classifier, we plotted model error rates. Figure 3 shows the kinds of errors that the decision tree, random forest, and support vector machine made. Rows represent actual classes and columns represent predicted classes. Across all models, social, constructive, and interactive engagement were less likely to be misclassified with each other. For the model trained with SVM, active and constructive as well as the active and interactive engagement levels were less likely to be misclassified with one another.

## 6. DISCUSSION

The purpose of this study was to develop models that identify cognitive engagement in online discussion forums. Using the post features that were extracted with Coh-Metrix, we trained three classifiers. Employing transparent classifiers allowed us to more easily evaluate feature importance for cognitive engagement prediction. We also conducted an error analysis to better understand the classification errors across models. Below, we discuss the implications of our findings in the context of each research question.

### 6.1 Feasibility of Automating the Prediction of Cognitive Engagement

Of the three types of predictive models we trained (i.e., decision tree, random forest, and support vector machine), the support vector machine performed sufficiently well for it to be used to identify engagement levels in discussion posts.

Previous studies generally focused on cognitive engagement prediction using the random forest algorithm and the CoI framework. We used a different cognitive engagement framework (i.e., ICAP), and our best model (SVM) demonstrated similar accuracy (70.3%) and Cohen’s kappa values ( $K = .63$ ) to those that used random forest under the CoI model of cognitive presence (e.g., [38, 59]).

The development of these models is the first step towards supporting instructors who want to be able to identify student engagement so that they can improve their teaching practices and intervene when student engagement levels are

low. These types of analytics fall into the category of those desired by instructors as they capture aspects of student post quality that are not currently available in most dashboards [1]. The models could be used to inform instructors about the engagement status of students in online course discussions. Given model performance, it would be important to effectively communicate uncertainty in its labelling of student posts [24]. Approaches, such as those suggested by Brooks and Greer [11], could be used to mitigate the risk of instructors relying on misclassified data so that they appropriately trust the output of the model and act in accordance with its limitations. For example, the system could identify the students who are disengaging to warn instructors so that they adjust their teaching and instructional practices. Additionally, such systems could be used to nudge students to better engage with tasks or to recommend posts that may enhance student engagement levels [14].

### 6.2 Feature Importance for the Prediction of Cognitive Engagement

By investigating how different features contribute to model performance, we can better understand the underlying phenomena which, in turn, will support the development of better predictive models for detecting cognitive engagement. Our study also provided further empirical evidence of feature importance, corroborating the findings of previous studies.

Both the decision tree and random forest feature importance analyses identified similar features (e.g., AWL count, Flesch-Kincaid grade level, word count). The features identified for the models in the present study were consistent with previous studies (e.g., [38, 59]). Similar to these previous studies, we found that higher levels of cognitive engagement were associated with longer messages. Our work builds on this by highlighting the importance of vocabulary use through the AWL count feature, which was not included in the other studies. Rather, those studies captured vocabulary through other means (e.g., Linguistic Inquiry and Word Count [84]). Our findings indicate that the use of academic vocabulary, like those in the AWL, that are not specific to a discipline (words such as hypothesis, conclude) support the identification of cognitive engagement. Moreover, their more general nature means that they should support generalization across courses at similar academic levels.

### 6.3 Interpreting Classification Errors

Classification errors occurred between the active and social engagement labels. Across all of the algorithms we used in our study, the worst performance was observed for active engagement. In a recent study, Farrow and colleagues [38] also found similar results for their prediction of active

engagement using random forest and the CoI framework. However, they did not perform an error analysis.

Our analysis of active and social engagement classes suggested that median word count, median AWL count, and Flesch-Kincaid grade level were similar for these two engagement levels. Furthermore, the dispersion of predictor values for active engagement was wide, overlapping with the range of predictors of other engagement levels. This helps explain what may have contributed to the poor performance of models for active engagement identification. Constructive engagement, on the other hand, had more distinct dispersion than other engagement levels, which could explain why the models were more successful when predicting constructive engagement. These results suggest the importance of having distinct engagement categories for successfully differentiating the vector space, thus achieving higher model performance and accuracy across engagement categories.

## 6.4 Limitations and Future Directions

Like all models, ours capture certain characteristics of student engagement and the decisions we made influence both their accuracy and the extent to which they are expected to generalize to other settings. We discuss these issues below and their potential for creating new opportunities for the development and use of models in online-learning settings.

The overall performance of the models suggests there is room for improvement. Nonetheless, these models can provide a snapshot of the engagement level in each post, providing an early signal of student engagement in online learning environments. This signal could help instructors to derive insight from this student-engagement data.

We labelled the data with the highest engagement level observed in a post. However, different cognitive engagement levels (e.g., social and constructive) may co-exist in a single post. While this type of coding approach is commonly employed [90], it may have limited the performance of our classifiers. It also prevents a more nuanced understanding of the types of cognitive engagement demonstrated by students. To support more nuanced representations of student engagement, future research can consider the co-existence of different engagement levels or how engagement relates to different areas of a course using something akin to aspect-based sentiment analysis [96].

Another limitation of our study is that we focused on identifying cognitive engagement in online discussion posts. That is, posts are classified rather than students. Because engagement may fluctuate based on course content or a weekly basis, we employed post-level analyses, which is a common practice in such research (e.g., [38, 70]). However, this may influence performance metrics. Future research can focus on how to derive an appropriate measure and representation of a student's cognitive engagement based on the varied levels of engagement that are exhibited across their posts. Perhaps more importantly, students who did not post to the discussion forums or those who only read posts and did not create messages (i.e., online listeners [92]) have not been included because there were no posts from these students. Alternative mechanisms need to be found to characterize their engagement through a listening lens. Since simply opening

a post is not enough to infer cognitive engagement, proxy measures may need to be developed. Behavioural patterns such as the ratio between the re-reading of posts and posts made [1, 72] or scan-rate [49] may provide reasonable signals of online listener engagement.

## 7. CONCLUSION

In this study, we gauged the feasibility of automating cognitive engagement prediction through NLP and ML methods. We first manually coded over 4,000 posts using the coding scheme we adapted from ICAP and Bloom's taxonomy. Then, we extracted linguistic and contextual features and trained three machine learning classifiers to predict the level of cognitive engagement demonstrated in a forum post. We obtained promising results with a support vector machine.

Now that we have models that can identify the highest level of cognitive engagement seen in a single post, we can start to consider how other factors might interact with student cognitive engagement. For example, we know that the course structure, length, and facilitation method influence student participation [72], language use [27], and social support [26, 72]. Additionally, the system used to deliver a course can influence student engagement [28, 79]. Future research could improve cognitive engagement identification by integrating such information.

This work makes several contributions, with some at the theoretical level and others at the empirical level. First, we mapped cognitive engagement levels onto Bloom's taxonomy. Bloom's taxonomy places cognitive complexity in hierarchical order, as does our ICAP-based cognitive engagement coding scheme. Second, this study illustrated the utility of different classifiers for cognitive engagement prediction in graduate-level online courses. Our analysis of the importance of model features and errors is consistent with previous studies; it confirmed the importance of word count for predicting cognitive engagement and revealed the importance of AWL count and Flesch-Kincaid grade level for predicting cognitive engagement. Future studies can include these features and use a support vector machine to develop a predictive model for cognitive engagement in online learning environments. Building on this work will enable the development of better models that can then be used to inform teaching and learning when online discussion forums are part of course delivery.

## 8. ACKNOWLEDGEMENTS

This work is supported in part by funding from the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2018-03834].

## 9. REFERENCES

- [1] G. Akcayir, L. Farias Wanderley, C. Demmans Epp, J. Hewitt, and A. Mahmoudi-Nejad. Learning analytics dashboard use in online courses: Why and how instructors interpret discussion data. In M. Sahin and D. Ifenthaler, editors, *Visualizations and Dashboards for Learning Analytics*, pages 371–397. Springer International Publishing, Cham, 2021. Series Title: Advances in Analytics for Learning and Teaching.

- [2] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, M. Munoz-Organero, et al. Delving into participants' profiles and use of social tools in moocs. *IEEE Transactions on Learning Technologies*, 7(3):260–266, 2014.
- [3] K. L. Alexander, D. R. Entwisle, and C. S. Horsey. From first grade forward: Early foundations of high school dropout. *Sociology of Education*, 70:87–107, 1997.
- [4] L. W. Anderson and D. R. Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, 2001.
- [5] J. J. Appleton, S. L. Christenson, D. Kim, and A. L. Reschly. Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, 44(5):427–445, 2006.
- [6] T. Atapattu, M. Thilakarathne, R. Vivian, and K. Falkner. Detecting cognitive engagement using word embeddings within an online teacher professional development community. *Computers & Education*, 140:1–14, 2019.
- [7] R. S. Baker, A. T. Corbett, I. Roll, and K. R. Koedinger. Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction (UMUAI)*, 18(3):287–314, Aug. 2008.
- [8] J. E. Beck. Engagement tracing: Using response times to model student disengagement. In C.-K. Looi, G. McCalla, B. Bredeweg, and J. Breuker, editors, *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*, pages 88–95, Amsterdam, The Netherlands, 2005. IOS Press.
- [9] G. Bonet and B. R. Walters. High impact practices: Student engagement and retention. *College Student Journal*, 50(2):224–235, 2016.
- [10] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth, 1984.
- [11] C. Brooks and J. Greer. Explaining predictive models to learning specialists using personas. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, LAK '14*, page 26–30, New York, NY, USA, 2014. Association for Computing Machinery.
- [12] W. Cade, N. Dowell, A. Graesser, Y. Tausczik, and J. Pennebaker. Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In *Proceedings of the 7th International Conference on Educational Data Mining*, pages 399–400, 2014.
- [13] R. M. Carini, G. D. Kuh, and S. P. Klein. Student engagement and student learning: Testing the linkages. *Research in Higher Education*, 47(1):1–32, 2006.
- [14] Z. Chen and C. Demmans Epp. CSCLRec: Personalized recommendation of forum posts to support socio-collaborative learning. In A. N. Rafferty, J. Whitehill, V. Cavalli-Sforza, and C. Romero, editors, *Thirteenth International Conference on Educational Data Mining (EDM 2020)*, pages 364–373, Fully Virtual, 2020. International Educational Data Mining Society.
- [15] M. T. Chi and R. Wylie. The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4):219–243, 2014.
- [16] M. Cocea and S. Weibelzahl. Log file analysis for disengagement detection in e-learning environments. *User Modeling and User-Adapted Interaction*, 19(4):341–385, 2009.
- [17] M. Cocea and S. Weibelzahl. Disengagement detection in online learning: Validation studies and perspectives. *IEEE transactions on learning technologies*, 4(2):114–124, 2010.
- [18] K. S. Cooper. Eliciting engagement in the high school classroom: A mixed-methods examination of teaching practices. *American Educational Research Journal*, 51(2):363–402, 2014.
- [19] L. Corno and E. B. Mandinach. The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, 18(2):88–108, 1983.
- [20] A. Coxhead. A new academic word list. *TESOL Quarterly*, 34(2):213–238, 2000.
- [21] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2(Dec):265–292, 2001.
- [22] P. A. Creed, J. Muller, and W. Patton. Leaving high school: The influence and consequences for psychological well-being and career-related confidence. *Journal of Adolescence*, 26(3):295–311, 2003.
- [23] J. DeBoer, A. D. Ho, G. S. Stump, and L. Breslow. Changing “course” reconceptualizing educational variables for massive open online courses. *Educational Researcher*, 43(2):74–84, 2014.
- [24] C. Demmans Epp and S. Bull. Uncertainty representation in visualizations of learning analytics for learners: Current approaches and opportunities. *IEEE Transactions on Learning Technologies*, 8(3):242–260, 2015.
- [25] C. Demmans Epp and K. Phirangee. Exploring mobile tool integration: Design activities carefully or students may not learn. *Contemporary Educational Psychology*, 59(2019):1–17, Oct. 2019.
- [26] C. Demmans Epp, K. Phirangee, and J. Hewitt. Student actions and community in online courses: The roles played by course length and facilitation method. *Online Learning*, 21(4):53–77, Dec. 2017.
- [27] C. Demmans Epp, K. Phirangee, and J. Hewitt. Talk with me: Student behaviours and pronoun use as indicators of discourse health across facilitation methods. *Journal of Learning Analytics*, 4(3):47–75, Dec. 2017.
- [28] C. Demmans Epp, K. Phirangee, J. Hewitt, and C. A. Perfetti. Learning management system and course influences on student actions and learning experiences. *Educational Technology, Research and Development (ETRD)*, 68(6):3263–3297, Dec. 2020.
- [29] C. Develotte, N. Guichon, and C. Vincent. The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, 22(3):293–312, 2010.

- [30] M. Dewan, M. Murshed, and F. Lin. Engagement detection in online learning: a review. *Smart Learning Environments*, 6(1):1–20, 2019.
- [31] N. Diana, M. Eagle, J. Stamper, S. Grover, M. Bienkowski, and S. Basu. An instructor dashboard for real-time analytics in interactive programming assignments. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 272–279, 2017.
- [32] N. M. Dowell, A. C. Graesser, and Z. Cai. Language and discourse analysis with coh-matrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, 3(3):72–95, 2016.
- [33] N. M. M. Dowell and A. C. Graesser. Modeling learners’ cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, 1(3):183–186, 2014.
- [34] J. J. Duderstadt, D. E. Atkins, D. E. Van Houweling, and D. Van Houweling. *Higher education in the digital age: Technology issues and strategies for American colleges and universities*. Greenwood Publishing Group, 2002.
- [35] S. D’Mello, B. Lehman, R. Pekrun, and A. Graesser. Confusion can be beneficial for learning. *Learning and Instruction*, 29:153–170, 2014.
- [36] J. S. Eccles, C. Midgley, A. Wigfield, C. M. Buchanan, D. Reuman, C. Flanagan, and D. Mac Iver. Development during adolescence: The impact of stage–environment fit on young adolescents’ experiences in schools and in families (1993). *American Psychologist*, 48(2):90–101, 1993.
- [37] E. Farrow, J. Moore, and D. Gašević. Analysing discussion forum data: a replication study avoiding data contamination. In *Proceedings of the 9th international conference on learning analytics & knowledge*, pages 170–179, 2019.
- [38] E. Farrow, J. Moore, and D. Gašević. Dialogue attributes that inform depth and quality of participation in course discussion forums. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 129–134, 2020.
- [39] R. Ferguson and D. Clow. Examining engagement: analysing learner subpopulations in massive open online courses (moocs). In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 51–58, 2015.
- [40] M. Fernández Delgado, E. Cernadas García, S. Barro Ameneiro, and D. G. Amorim. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15(1):3133–3181, 1993.
- [41] J. D. Finn and D. A. Rock. Academic success among students at risk for school failure. *Journal of Applied Psychology*, 82(2):221–234, 1997.
- [42] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1):59–109, 2004.
- [43] J. A. Fredricks and W. McColskey. The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, and C. Wylie, editors, *Handbook of research on student engagement*, pages 763–782. Springer, Boston, MA, USA, 2012.
- [44] S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, 111(23):8410–8415, 2014.
- [45] K. A. Fuller, N. S. Karunaratne, S. Naidu, B. Exintaris, J. L. Short, M. D. Wolcott, S. Singleton, and P. J. White. Development of a self-report instrument for measuring in-class student engagement reveals that pretending to engage is a significant unrecognized problem. *PloS one*, 13(10):1–22, 2018.
- [46] D. R. Garrison, T. Anderson, and W. Archer. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3):87–105, 1999.
- [47] A. C. Graesser, D. S. McNamara, M. M. Louwerse, and Z. Cai. Coh-matrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2):193–202, 2004.
- [48] H. Han, X. Guo, and H. Yu. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 219–224. IEEE, 2016.
- [49] J. Hewitt, C. Brett, and V. Peters. Scan Rate: A New Metric for the Analysis of Reading Behaviors in Asynchronous Computer Conferencing Environments. *American Journal of Distance Education*, 21(4):215–231, Nov. 2007.
- [50] M. Hu and H. Li. Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)*, pages 39–43. IEEE, 2017.
- [51] J.-L. Hung and K. Zhang. Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*, 4(4):426–437, 2008.
- [52] H. Iraj, A. Fudge, H. Khan, M. Faulkner, A. Pardo, and V. Kovanović. Narrowing the feedback gap: Examining student engagement with personalized and actionable feedback messages. *Journal of Learning Analytics*, 8(3):101–116, 2021.
- [53] M. Jacobi et al. *College Student Outcomes Assessment: A Talent Development Perspective*. ASHE-ERIC Higher Education Report No. 7. ERIC, 1987.
- [54] J. Johns and B. P. Woolf. A dynamic mixture model to detect student motivation and proficiency. In *21<sup>st</sup> National Conference on Artificial Intelligence (AAAI)*, pages 163–168, Boston, MA, USA, 2006.
- [55] A. Kapoor and R. W. Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682, 2005.
- [56] S. N. Karimah and S. Hasegawa. Automatic engagement recognition for distance learning systems: A literature study of engagement datasets and methods. In *International Conference on*

- Human-Computer Interaction*, pages 264–276. Springer, 2021.
- [57] R. F. Kizilcec, C. Piech, and E. Schneider. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 170–179, 2013.
- [58] V. Kovanović, S. Joksimović, D. Gašević, and M. Hatala. Automated cognitive presence detection in online discussion transcripts. In *CEUR Workshop Proceedings*, volume 1137, pages 1–4, 2014.
- [59] V. Kovanović, S. Joksimović, Z. Waters, D. Gašević, K. Kitto, M. Hatala, and G. Siemens. Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the sixth international conference on learning analytics & knowledge*, pages 15–24, 2016.
- [60] G. D. Kuh. *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties*. Indiana University Center for Postsecondary Research, 2001.
- [61] G. D. Kuh. The national survey of student engagement: Conceptual and empirical foundations. *New directions for institutional research*, 141:5–20, 2009.
- [62] G. D. Kuh, T. M. Cruce, R. Shoup, J. Kinzie, and R. M. Gonyea. Unmasking the effects of student engagement on first-year college grades and persistence. *The journal of higher education*, 79(5):540–563, 2008.
- [63] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174, 1977.
- [64] E. S. Lane and S. E. Harris. A new tool for measuring student behavioral engagement in large university classes. *Journal of College Science Teaching*, 44(6):83–91, 2015.
- [65] S. Li, S. P. Lajoie, J. Zheng, H. Wu, and H. Cheng. Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving. *Computers & Education*, 163:1–13, 2021.
- [66] B. D. Lutz, A. J. Barlow, N. Hunsu, C. J. Groen, S. A. Brown, O. Adesope, and D. R. Simmons. Measuring engineering students’ in-class cognitive engagement: Survey development informed by contemporary educational theories. In *2018 ASEE Annual Conference & Exposition*, pages 1–15, 2018.
- [67] L. P. Macfadyen and S. Dawson. Mining lms data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2):588–599, 2010.
- [68] I. Molenaar and C. A. Knoop-van Campen. How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies*, 12(3):347–355, 2018.
- [69] B. A. Motz, J. D. Quick, J. A. Wernert, and T. A. Miles. A pandemic of busywork: Increased online coursework following the transition to remote instruction is associated with reduced academic achievement. *Online Learning*, 25(1):70–85, 2021.
- [70] V. Neto, V. Rolim, A. Pinheiro, R. D. Lins, D. Gašević, and R. F. Mello. Automatic content analysis of online discussions for cognitive presence: A study of the generalizability across educational contexts. *IEEE Transactions on Learning Technologies*, 14(3):299–312, 2021.
- [71] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [72] K. Phirangee, C. Demmans Epp, and J. Hewitt. Exploring the relationships between facilitation methods, students’ sense of community and their online behaviours. *Special Issue on Online Learning Analytics. Online Learning Journal*, 20(2):134–154, 2016.
- [73] J. Pietarinen, T. Soini, and K. Pyhältö. Students’ emotional and cognitive engagement as the determinants of well-being and achievement in school. *International Journal of Educational Research*, 67:40–51, 2014.
- [74] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor. Uncovering hidden engagement patterns for predicting learner performance in moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 157–158, 2014.
- [75] S. Raschka. Mlxtend: Providing machine learning and data science utilities and extensions to python’s scientific computing stack. *The Journal of Open Source Software*, 3(24), Apr. 2018.
- [76] S. Raschka. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, 2018.
- [77] K. A. Renninger and J. E. Bachrach. Studying triggers for interest and engagement using observational methods. *Educational Psychologist*, 50(1):58–69, 2015.
- [78] C. Romero and S. Ventura. Educational Data Mining: a Survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135–146, July 2007.
- [79] C. P. Rosé and O. Ferschke. Technology support for discussion based learning: From computer supported collaborative learning to the future of massive open online courses. *International Journal of Artificial Intelligence in Education*, 26(2):660–678, June 2016.
- [80] M. Schonlau and R. Y. Zou. The random forest algorithm for statistical learning. *The Stata Journal*, 20(1):3–29, 2020.
- [81] D. J. Shernoff. *Optimal learning environments to promote student engagement*. Springer, 2013.
- [82] D. J. Shernoff, M. Csikszentmihalyi, B. Schneider, and E. S. Shernoff. *Student engagement in high school classrooms from the perspective of flow theory*, pages 475–494. Springer, 2014.
- [83] D. J. Stipek. *Motivation to learn: Integrating theory and practice*. Pearson College Division, 2002.
- [84] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.
- [85] S.-F. Tseng, Y.-W. Tsao, L.-C. Yu, C.-L. Chan, and K. R. Lai. Who will pass? analyzing learner behaviors

in moocs. *Research and Practice in Technology Enhanced Learning*, 11(1):1–11, 2016.

- [86] C. O. Walker and B. A. Greene. The relations between student motivational beliefs and cognitive engagement in high school. *The Journal of Educational Research*, 102(6):463–472, 2009.
- [87] C. O. Walker, B. A. Greene, and R. A. Mansell. Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, 16(1):1–12, 2006.
- [88] N. Wanas, M. El-Saban, H. Ashour, and W. Ammar. Automatic scoring of online discussion posts. In *Proceedings of the 2nd ACM workshop on Information credibility on the web*, WICOW '08, pages 19–26, New York, NY, USA, 2008. ACM.
- [89] M.-T. Wang and R. Holcombe. Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, 47(3):633–662, 2010.
- [90] X. Wang, M. Wen, and C. P. Rosé. Towards triggering higher-order thinking behaviors in moocs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 398–407, 2016.
- [91] J. Whitehill, M. Bartlett, and J. Movellan. Automatic facial expression recognition for intelligent tutoring systems. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6. IEEE, 2008.
- [92] A. F. Wise, S. N. Hausknecht, and Y. Zhao. Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*, 9(2):185–209, June 2014.
- [93] J.-Y. Wu, J. N. Hughes, and O.-M. Kwok. Teacher–student relationship quality type in elementary grades: Effects on trajectories for achievement and engagement. *Journal of School Psychology*, 48(5):357–387, 2010.
- [94] E. Yazzie-Mintz and K. McCormick. *Finding the humanity in the data: Understanding, measuring, and strengthening student engagement*, pages 743–761. Springer, 2012.
- [95] E. Yogev, K. Gal, D. Karger, M. T. Facciotti, and M. Igo. Classifying and visualizing students' cognitive engagement in course readings. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pages 1–10, 2018.
- [96] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges, 2022.
- [97] Y. Zhao, S. Lin, J. Liu, J. Zhang, and Q. Yu. Learning contextual factors, student engagement, and problem-solving skills: A chinese perspective. *Social Behavior and Personality: an international journal*, 49(2):1–18, 2021.
- [98] E. Zhu. Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science*, 34(6):451–480, 2006.