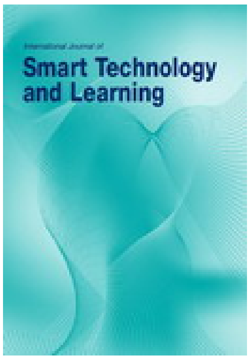
International Journal of Smart Technology and Learning

ISSN online: 2056-4058 - ISSN print: 2056-404X
https://www.inderscience.com/ijsmarttl



Task-agnostic team competence assessment and metacognitive feedback for transparent project-based learning in data science

Hong Liu, Timothy Bernard, Elif Cankaya, Alex Hall

DOI: 10.1504/IJSMARTTL.2023.10054696

Article History:

Received:	25 April 2021
Last revised:	31 December 2021
Accepted:	08 April 2022
Published online:	17 March 2023

Task-agnostic team competence assessment and metacognitive feedback for transparent project-based learning in data science

Hong Liu*, Timothy Bernard, Elif Cankaya and Alex Hall

Embry-Riddle Aeronautical University,
Daytona Beach, FL 32114, USA
Email: liuho@erau.edu
Email: Darby.bernard@gmail.com
Email: emctxoh@gmail.com
Email: Halla22000@gmail.com
*Corresponding author

Abstract: Assessing team and individual competencies from team projects' outcomes alone can be pretty subjective. Sharing credit for team efforts equally between team members or differentiating individual contributions based on peer evaluations that might be prone to bias destroys motivation and hinders learning. A fair assessment of individual performances should depend on a formative assessment of a team's process and each individual's contribution to tasks. Such an assessment is time-consuming and only affordable to utilise in small classes. This research serves as a small step to synergise the human and Artificial Intelligence (AI) based educational technology to improve the transparency and effectiveness of collaborative Project-Based Learning (PBL). We introduce a web-bot (BotCaptain) to automate parts of the instructional tasks, present a task-agnostic team competency model, and recommends a set of metacognitive feedback for team members. Study findings have implications for the use of AI in PBL environments.

Keywords: teamwork; project based learning; computer supported collaborative learning; competency based learning assessment; metacognitive feedback; artificial intelligence.

Reference to this paper should be made as follows: Liu, H., Bernard, T., Cankaya, E. and Hall, A. (2023) 'Task-agnostic team competence assessment and metacognitive feedback for transparent project-based learning in data science', *Int. J. Smart Technology and Learning*, Vol. 3, No. 2, pp.138–162.

Biographical notes: Hong Liu is a professor in Data Science at Embry-Riddle Aeronautical University, USA. His recent research has focused on applying Artificial Intelligence technology and Learning Analytics to promote collaborative learning in a hybrid learning environment.

Timothy Bernard is currently a master's student in the Data Science program at Embry-Riddle Aeronautical University, USA. He served as the software development team leader for the project associated with this research paper 2020–2022. He has been hired by IBM, starting to work in June 2022.

Elif Cankaya is currently a master's student in the Data Science program at Embry-Riddle Aeronautical University, USA. She holds a PhD in Educational Psychology and is currently working for DISH Network as Anaplan Analyst/ Developer and part of company's CEO team.

Alex Hall is currently a master's student in the Data Science program at Embry-Riddle Aeronautical University. He has been hired as a Data Engineer by Space-Eyes. He will be architecting the data platform to ingest satellite and other sensor data for machine learning applications starting in June 2022.

1 Introduction

Team projects are great opportunities for learners to gain skills and develop competencies, including collaboration skills and experience that are highly valued in almost all professional work environments. Given the importance of teamwork in today's competitive workspace, ensuring accurate evaluation of each individual learner's performance and contribution, and delivering effective and timely feedback to the learners over the course of the project are vital. However, when assessing team projects, the traditional assessment method is mainly based on peer evaluations and the instructor's evaluation of the final project products (i.e., presentations, reports, etc.). While peer evaluations may prone to bias, assigning the same or relatively similar credits to each team member based on project products without considering each team member's individual effort and performance is problematic. This shared grading approach is not desirable, especially if the team projects constitute a large percentage of the course grade; such a grading method is unfair and may destroy learners' motivation.

Ideally, the instructor should assess the whole PBL process that may span from weeks to months, not only the final products. Moreover, grading should reflect each team member's own contribution to the team, role on task and performance. Like team sports where each individual athlete's contribution and performance are closely tracked by the coaches' and remain under fans' scrutiny, carefully monitoring each team member's progress in PBL is necessary to establish transparency and eliminate potential bias in assessment. Also, like the way that coaches deliver feedback to each athlete based on individual progress and the unique roles in team; PBL should involve ongoing and timely feedback throughout the project to each team member to ensure an effective PBL environment.

Inspired by team sports, the authors of the study opine that PBL as a teaching method should rely on best practices used in team sport to ensure the successful implementation of PBL into academic courses. Similar to coaching sport teams, instructors should closely monitor each team member's progress, acknowledge every single team member's contribution through fair performance grading, and provide prompt individualised feedback. However, such an approach is costly due to limited allocated instructor time and resources since monitoring an average class size of students through the course of the PBL requires a great deal of time, not to mention the popular Massive Open Online Courses (MOOCs) with quite large student enrolment.

Considering the scarcity of time and resources, Artificial Intelligence (AI) technologies could be utilised to assist instructors for ensuring a fair performance assessment that truly reflects the individual effort and contribution and providing timely feedback to all learners (Spector and Wu, 2017). The use of AI in the implementation of PBL is the main motivation for this research that came from the need for the Master's Degree of Science in Data Science (MSDS) program offered at the Embry-Riddle Aeronautical University. Data Science (DS) is an emerging interdisciplinary domain that uses machine learning and statistics to derive insights from available data to support or optimise decisions in a wide range of application domains (i.e., finance, health care, education). Solving a real-world problem in DS depends on teamwork due to its intrinsic interdisciplinary nature. Based on an active learning curriculum design, the MSDS courses heavily rely on PBL in Computer-Supported Collaborative Learning (CSCL) environment (Orvis and Lassiter, 2008). CSCL collects massive amounts of granular data and analyses them offline, which is slow, labour-intensive, and challenging to provide immediate feedback.

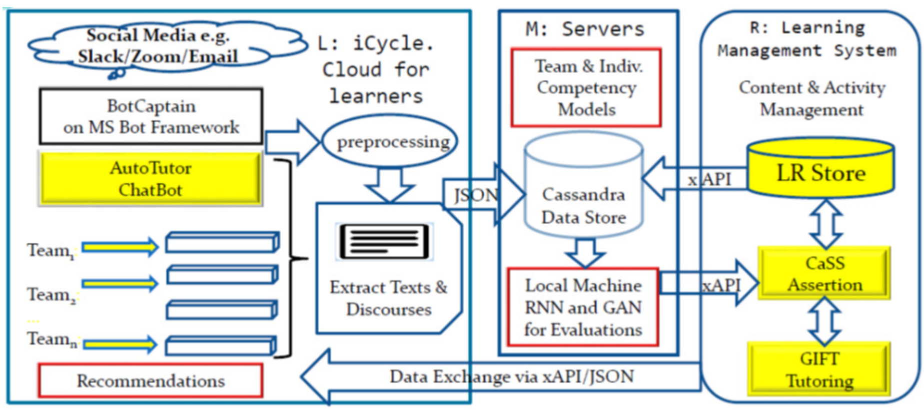
To improve the effective implementation of PBL in the MSDS program, this paper focuses on how to utilise AI to assist instructors with their instructional roles and duties for successfully managing PBL environments. More specifically, the purpose of the study is to assist instructors in dealing with their task-agnostics management duties that include collecting formative assessment data, monitoring communication between peers, and providing immediate feedback in their DS courses heavily based on PBL. Instructional domain-specific responsibilities such as answering discipline-specific and theoretical questions about DS concepts are too complex and challenging to automate at this stage. However, successful completion of DS projects depends not only on the team members' collective cognitive abilities and domain-specific proficiency, but also on team's non-cognitive traits and task-agnostic teamwork skill that aimed to be monitored and supported using AI technologies.

In this study, we designed and prototyped an intelligent computer-supported *hybrid* collaborative learning environment (iCycle, Liu et al., 2020) that aims to support instructors in dealing with their task agnostic duties so that PBL environment in MSDS courses can be transformed into exciting, engaging, and transparent sports-like games environments. As indicated in the above paragraph, at the current stage, the instructors and the TA are responsible for handling domain-specific duties such as answering all domain-specific questions about DS concepts and guiding learners to solve task-specific problems.

Figure 1 illustrates the revised System Conceptual Model and Components of iCycle. The major features of BotCaptain (https://n0m4d.gitbook.io/botcaptain/) include collecting communication data for tracking the interactions of teammates, coordinating team meetings, and monitoring the progress of teamwork. The current progress has focused on integrating BotCaptain with other tools by using Experience Application Program Interface (xAPI) interoperable data exchange standards and databases. The output data of the BotCaptain are stored in two cloud database systems – Cassandra in Microsoft Azure Cloud and Learning Record Store (LRS) in Amazon Web Services (AWS). Veracity Technology Consultants donated us the LRS system license (https://LRS/io). Temporary data are stored in Cassandra for the learning analytics (LA) extension to assess team competencies and provide task-agnostic metacognitive feedbacks. The more permanent data are stored in LRS with xAPI to exchange data with Moodle and potentially other tools such as the General Intelligence Framework for Tutoring (GIFT, see Sottilare et al., 2018). Through interoperable data links such as xAPI, the BotCaptain, its LA extension, and other third-party tools in shaded yellow boxes such as CaSS (Competency and Skill System) of EduWorks and AutoTutor developed by the Institute for Intelligent Systems of the University Memphis are integrated with low coupling. AI and BotCaptain used in iCycle, collect the necessary data to make teamwork transparent and provide each learner with timely feedback while protecting student privacy.

This article is organised into six sections. Section 1 includes the Introduction. Section 2 addresses the Content and Activity Management, including the Teamwork Process for a PBL in the DS course. The next three sections describe the three components marked by the three red boxes in Figure 1. More specifically, Section 3 discusses the Task-agnostic Team Competence (TATC) assessment, including individual competency assessment, emphasising the task-agnostic aspect of teamwork. We describe LA Extension in Section 4 and present the recommendations focusing on metacognitive feedback in Section 5. The last section provides a conclusion and addresses directions for future work.

Figure 1 The system conceptual model and components of iCycle



2 Team formation, team process, and learning activities

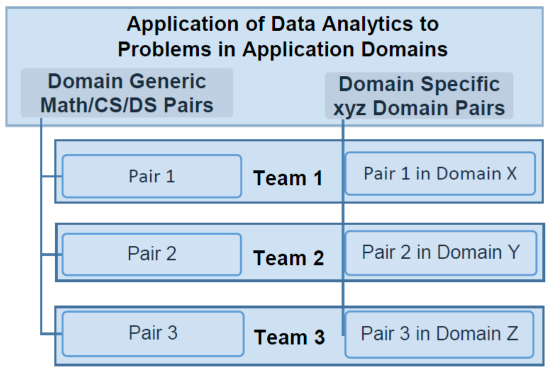
Team assessments include team outcomes, team processes, and observable behaviour markers that indicate good or poor teamwork (Messick, 1994). This section addresses the team process, including learning activities.

2.1 Team formation

The MSDS program at the Embry-Riddle Aeronautical University has six tracks to attract students from five different undergraduate disciplines: business, aviation safety, human factors, engineering, computer science, and mathematics. Students in the six tracks share five core data analytic courses – Statistics, Database, Data Visualisation, Data Mining, and Data-driven Modelling. They can flexibly choose five elective courses that fit their interested domain of applications. The diversity of student backgrounds becomes a strength for developing team projects that fit the interdisciplinary nature of DS. Except for Statistics, a more traditional course, the other four core courses all use PBL teamwork as one-third of the course during the last five weeks of the semester. Since the grade of the project counts for nearly half of the course grade, fair grading based on the student's performance, efforts, and contributions become critical.

A typical team consists of four students with complementary skills. Most students prefer to work with their close classmates or with students from the same undergraduate degree programs with similar academic background if given the option to choose their teammates. However, like project teams in the workplace, a team formed to complete a task needs to have adequate skills and collective expertise. For example, if a team is formed to use data analytics to solve an aviation safety problem, the team needs students with strong programming skills and students who have domain knowledge in aviation safety (see Liu et al., 2018 for details).

Figure 2 Team formation with pairs of friends with complementary skills



As shown in Figure 2, students are assigned into four-member teams. Each team consisted of two pairs of students from two different types of background. One pair has domain-generic skills in math modelling and programming, and the other pair has domain-specific knowledge and skills such as aviation safety or business. Next, the instructor asked students to choose a friend and propose sketchy project ideas. Then, the instructor evaluated the academic background of all pairs and their proposed project ideas to make sure that each team has members with complementary skills and knowledge to solve a mutually agreeable project. After proposed projects are approved, the instructor serves as research facilitators to help students through the data analysis and interpretation processes as well as to answer any technical questions in a timely manner. Specifically, the instructor has the following tasks: 1) Provide or approve an open-ended research project including the datasets; 2) Set project ground rules such as how often students should exchange ideas either face-to-face or online, how students should write up references, handle conflict, and so forth.; 3) Promote learner-learner interactions, and 4) Monitor progress and relationships while removing obstacles.

2.2 Team project process

According to Mathieu et al. (2008), the team project process is perhaps more important than project outcomes in a training setting. The team process addresses the members' dynamic roles in terms of tasks and member interactions, such as how information is shared and decisions are made. In the team projects in DS courses, the scores of team projects count for about 40% of the course grade. We consider that the team process is as important as the team outcomes. Therefore, 20% of the course score is assigned to the team outcomes that the team members are most likely to share equally. However, the other 20% designated to the team process are evaluated individually based on the TATC assessment with the details described in the next section. The two sources of input for assessing the team process are the metadata collected at weekly meetings with the instructor (the roles on tasks of each member every week) and the peer-evaluations.

After the teams are formed, the instructor needs to:

1 Provide a timeline for the project and rubrics about evaluating outcomes.

2 Prepare a shared folder for students to submit intermediate artefacts such as datasets, data cleansing, graphs, etc.

3 Meet the students weekly to get progress reports, update them about their roles on tasks, and help remove obstacles.

2.3 Team PBL related learning activities

A properly designed process and grading policy motivate students to learn collaboratively and work cooperatively (Olivares, 2008; Smith-Jentsch et al., 2008; Liu et al., 2018). While the team process is more independent of the content, the learning activities are content specific. For example, the learning activities in the Data Visualisation course differ from the Data Mining course. While the former emphasises exploration and qualitative explanation of data, and the latter requires quantifying the performance and uncertainty of predictive models. Nevertheless, the two courses share a similar team process regarding each member’s roles, and the ground rules of coordination, communication, and cooperation are the same. For a smooth and successful project process, the instructor sets in advance the project’s weekly milestones, with three to four tasks for each week. Two students are assigned to tackle each task, with one playing a leading role and the other playing a supporting role. Ideally, each student plays a leading role for one task and a supporting role for another task to ensure collaborative learning across the whole team. For example, if the leading programmer is in charge of developing the computing models, then the supporter is responsible for validating and testing the models. If the leading author writes the final report, the supporter must proofread and polish the writing. This design coerces collaborative learning and makes both accountable for the assigned task. Students are strongly encouraged to communicate with teamwork software such as Slack, MS Teams, or GroupMe for efficiency and transparent information exchange. Moreover, BotCaptain is a web-bot app that extends the teamwork software above so that the data from different platforms and LMS can be merged and analysed on our local server.

3 Traits and behaviour markers for TATC metadata

Team projects in DS and team sports share similar task-agnostic competence models. The individual competence models for both sports team and academic project team share motivational, affective, and interpersonal traits (Rosé et al., 2017). The team competence models for both sports teams and academic project teams share the four types of traits, including leadership/initiation, coordination, communication, and cooperation (Sottilare et al., 2018, 2020; Burmester, 2020; Biasutti and Frate, 2018). The TATC models we describe in this section answer the following four questions: (1) What inputs does the instructor provide? (2) What types of data for observable behavioural indicators are collected? (3) How do we collect the data? And (4) How to quantify the behavioural markers for the individual and collective non-cognitive traits?

3.1 Traits of TATC and the metadata for behavioural markers

The instructor provides each team with a project timeline and a shared cloud space for submitting the deliverables, weekly meeting minutes, and progress reports. Tables 1–4 illustrate the four facets of traits for the team competence model using the GIFT Team Meta-models (Sottilare et al., 2018, 2020; Burmester, 2020; McCormack et al., 2020). The behavioural markers include actions, spoken or written messages, and manners that help uncover the hidden antecedents of teamwork. Table 1 for team communication describes how the information is delivered. Table 2 for team cooperation, collaboration, and support describes attitudes and actions that help another team member with a task, such as anticipating a need for help and agreeing to help. Table 3 for team leadership and initiation describes the activities for guiding the team and initiating a critical task. Finally, Table 4 addresses information exchange that describes skills associated with maintaining shared situational awareness, such as knowing what information to share and with whom or understanding the team status and seeing the big picture. Some behaviour markers may intersect two facets of the traits. For example, negative emotion is a behaviour marker relating to a team's communication and collaboration traits.

Table 1 Team communication (data logged by BotCaptain)

Traits	*Behaviour markers*
+: Positive conversation showing mutual respect	Showing solidarity, agreeing, tension release, team-focused language (we, us), and showing listening, and empathy.
–: Negative conversation red-flagged Improper words	The occurrence of red-flagged words such as racial slur, sexist, profane, or other rude language. The offended students are encouraged to report such incidents.
N: Polarity of words in conversation	The ratio of positive and negative words in logged script of conversation
–: Rude or irrespective behaviours or conversations	Members constantly interrupt each other or talk in pairs without listening to the individual who has the floor.
N: deviation of speaking time/words	One or two members monopolise discussion throughout the meeting, and others rarely speak.
N: Promptness to respond to asynchronous messages	The average waiting time to respond to request of teammates, and the average waiting time to make the teammates aware of the situation and get helps.

All measurement of the data, including numerical data (N) such as deviation in speaking time, is transformed into five levels of ordinal data from 1 to 5 for easy processing by both humans and computers. For example, a positive trait scale ranging from poor, fair, good, very good, to excellent, and the data for a negative trait may have five levels based on the frequency of a behaviour ranging from never to often. Moreover, deviation of speaking time can be discretised into five equal bins such as most uneven, very uneven, average, reasonably even, and almost even. In the first column of the table, we use '+' and '–' signs for positive and negative qualitative data and N to label numerical data. Many qualitative traits, such as those in Table 4 are relatively subjective even when we count the frequency of occurrence of the behavioural markers. When the data are collected, the instructor can mitigate possible bias by comparing them with markers collected from other teams in the same or previously taught classes.

Table 2 Team cooperation, collaboration, and support

Traits	*Behaviour markers*	*Data Sources*
N: Unfairly assigned duties	The same individual or individuals end up doing the majority of the work, while other members either passively participate in or dictate.	Progress report and meeting minutes
N: Frequently delay meeting	Team meetings generally begin 5–15 minutes late, members often arrive late, leave early, or never even show up for the meetings.	Meeting Minutes BotCaptain logs
–: Negative emotion	With words or by appearance, some members clearly convey that they would rather be elsewhere.	Scripts of meetings Conversation
–: Consistent procrastination	One or more issues do not get resolved, only put on the back burner until next time.	Progress report BotCaptain logs
–: Lacking follow-up	No follow-up action plan is developed. Members are confused with regard to what the next step is and who is responsible for performing it.	Progress report BotCaptain logs

Table 3 Team leadership and initiation

Traits	*Behaviour markers*	*Data Sources*
–: Lacking a team manager	The team fails to elect a team manager – neither one serves for the whole project, nor the members turn to serve the role.	Meeting Minutes Progress report
–: Lacking visions	The team lacks a common vision, or has a vague definition of success, none or few members can articulate what is to accomplish.	Report to the instructor
–: No leader for critical task	The team cannot identify a leader and a supporter for some tasks. This causes the team misses deadlines or fails for critical components.	Meeting Minutes Progress report
–: Passiveness, lack initiative	No or few members take initiative to do research, depending on the instructor to a give research problem or spoon-feed technical solutions	lack of/ interaction with the instructor

3.2 *Traits of individual competence for teamwork and the metadata types for behavioural markers*

The literature typically divided non-cognitive personal traits into motivational, affective, and interpersonally dimensions (Rosé et al., 2017, Brenna et al., 2018). Given that leadership play a vital role in successful teamwork, leadership skillset was added as the fourth individual competence trait. Consequently, the non-cognitive personal traits and behavioural markers contributing to good or poor teamwork are categorised into the four following facets: (1) motivational, (2) affective, (3) interpersonal and social trait, and (4) leadership and initiations.

Table 4 Information exchange

Traits	*Behaviour markers*	*Data sources*
–: Confusing roles	Two or more members are unclear about weekly roles of team manager, technical leaders, or support roles.	Meeting Minutes Progress report
–: No Meeting agenda	Meeting lacks clear agenda, issues to resolve, and tasks to complete, the members simply have a vague notion of what they want to do.	Meeting Agenda and Minutes
–: Lack written commitments	The meetings run on and on and on with little to show for the time spent on them. Some tasks are left without anyone to be in charge.	Meeting Minutes Progress report
–: Fail to submit intermediate artefacts	The instructor sets a timeline for submitting intermediate artefacts. A team lacks leadership may fails to submit the intermediate artefacts and report progress, nobody knows who is in charge of what.	Artefacts submitted in shared folders and progress report
–: Unclear deadlines	Two or more are unclear the deadlines of deliverable, resulting in missed deadlines for required weekly deliverables.	Progress report BotCaptain logs

Table 5 Motivational trait of team member and behaviour markers

Traits	*Behaviour markers*	*Data Sources*
+: Accountability	The fulfilled duties and timeliness, in case of need, the effort to ask help from teammates or instructor, or TA.	Meeting Minutes Progress report
+: Effort	The work ethic, the hours on the task, and effort to make teammate be aware of this progress	Meeting Minutes & Peer evaluation
+: Grit	The persistence, and endurance to work under pressure.	Peer evaluation Role on tasks
+: Preparedness or readiness	Members have read the assignment, performed the necessary background research, or do what they were expected to do. Consequently, individuals are prepared for the meeting.	Meeting Minutes & Peer evaluation
–: Procrastination	The frequency of missing the due time for committed tasks and need for teammate to push to meet deadlines.	Progress report Communication
N: Promptness to respond to messages	The average waiting time to respond to request of teammates, and to make the teammates aware of the situation, and provide helps.	BotCaptain Logs

Table 6 Affective traits and behaviour markers of team member (data logged by BotCaptain)

Trait	*Behaviour markers*
+: Attitude	Showing solidarity, team-focused language (we, us), and showing listening and empathy.
+: Proactive	Take initiative to discuss issues, encourage teammate to brainstorm, integrate ideas, and resolve issues.
N: Polarity of words in conversation	The ratio of positive and negative words in logged script of conversation
N: Red-flagged Improper words	The occurrence and frequency of red-flagged words such as racial slur, sexist, profane, or other rude language.
–: Emotional stability	The frequency of anger issues or improper tones in teamwork when there is sharp disagreement.

Table 7 Interpersonal and social traits and behaviour markers (data logged by BotCaptain)

Traits	*Behaviour markers*
+: Speech skills	Active listen and take the proportional time to speak out. Avoid lowest common denominator between all points of view
+: Persuasiveness	Doing research to come up with evidence, factors not just opinions, Showing the pros and cons of alternative ideas.
+: Share credit & encouraging	building on the others' ideas and sharing credit fairly
+: Empathy and understanding	Showing agreeing and approval to others frequently, Tension Release, Agreeing,
+: Response to constructive criticism	Not take issue personal, and distinguish what matters to the project and seek resolution with positive attitude and advice

Table 8 Team member's leadership and initiation

Traits	*Behaviour markers*	*Data sources*
+/– Initiative	The member triggers more or less than the mean number of new ideas and discussions the project	BotCaptain logs & Meeting Minutes
+/– Integrations	The member makes more or less than the mean number of integrations other's ideas	BotCaptain logs & Meeting Minutes
+/– Resolution	The member makes more or less than the mean number of the resolution of discussions	BotCaptain logs & Meeting Minutes
+/– Win more or less supports	The member wins more or less than the mean numbers of supports from peers for his or her initiation	BotCaptain logs & Meeting Minutes
+/– Technical leadership	Take one of more major technical components such as programming, and develop mathematical models, writing, and literature review	Meeting Minutes and final report

3.3 The data collection mechanism and metrics for team competence assessment

An advantage of working in team projects in CSCL platforms, particularly on iCycle that includes the BotCaptain to empower data collection, is the rich and complete data sources as a by-product of collaborative work. The third column of six tables above indicates the data sources for the behavioural markers. As reflected, there are three origins of the data. The first is the synchronous Zoom meetings, and the second is the asynchronous communications through email, Slack, MS Teams, or other social media for teamwork. The third source is the submitted meeting minutes, progress reports, documents, programs, and dataset from the shared Google drive that the instructor prepared for each team. Whether the conversation is oral or written, the messages are transformed into texts. Details on functionality and data services can be found in the aforementioned Gitbook of BotCaptain.

Other data sources to validate the TATC assessment are the peer evaluation of the behaviours of teammates (see Table 9) and comprehensive peer evaluation (see Table 10) (Oakley et al., 2004). Due to the summative and comprehensive nature of the peer evaluation, the scale of Table 10 is more granular than the five levels of formative assessment data in Tables 1–8. These assessment forms are collected at the end of the project. Cross-validation of the team trait evaluation above with the peer evaluation data and individual trait evaluation data is used to mitigate any possible bias. In case the data from three sources do not yield consistent results, we provide the benefit of the doubt to the learner.

Table 9 Peer evaluation for teammates

	Team behaviour indicators	*Poor*	*Marginal*	*Fair*	*Good*	*Excellent*
1	Has the student attended team meetings?					
2	Has the student made a serious effort at assigned work before the team meetings?					
3	Has the student made a serious effort to fulfil his/her team role responsibilities on assignments?					
4	Has the student notified a teammate if he/she would not be able to attend a meeting or fulfil his/her tasks?					
5	Does the student communicate in group meetings or offline professionally and respectfully?					
6	Does the student listen to his/her teammates' ideas and opinions respectfully and give them careful consideration?					
7	Does the student cooperate with the group effort?					

Designed for human inputs, the data in Tables 9 and 10 are more subjective and summative than the other tables acquired from computer logs and used for AI-based assessment. Peer evaluation often reflects more on the personal relationship than the actual contributions. Nevertheless, the instructor can estimate who did the most and who did the least based on the collective inputs from all students. The datasets of all ten tables can provide a holistic assessment of the team process, enabling the instructor to evaluate individual contributions and efforts more accurately and fairly than just evaluating the final team product.

Table 10 Comprehensive peer evaluation to teammates

Excellent	Consistently carried more than his/her fair share of the workload and communicate timely, professionally, and respectfully
Very good	Consistently did what he/she was supposed to do, very well prepared and cooperative, and communicate with teammate timely and respectfully
Satisfactory	Usually did what he/she was supposed to do, acceptably prepared and cooperative, and communicate with teammate mostly timely and respectfully
Medium	Often did what he/she was supposed to do, minimally prepared and cooperative
Marginal	Sometimes failed to show up or complete assignments, rarely prepared
Deficient	Often failed to show up or complete assignments, rarely prepared
Unsatisfactory	Consistently failed to show up or complete assignments, rarely prepared
Superficial	Practically no participation
No show	No participation at all

4 The educational data mining and LA extension

A natural language processing (NPL) and Recurrent Neural Network (RNN) pipeline for LA has been developed for processing big educational data using SciKit Learning Python Packages. However, challenges to the use of behavioural markers in LA to evaluate student non-cognitive competencies came from three factors:

1. The lack of a data set that is large enough so far for reliable machine learning.
2. The data attributes are imbalanced and contain far more positive data attributes than negative data attributes.
3. The psychological phenomenon of reactivity occurs when individuals alter their performance or Behaviour due to the awareness that they are being observed (wiki/Reactivity_(psychology)).

The MSDS at the Embry-Riddle Aeronautical University is a small program, and our classes rarely exceed sixteen students (four teams). Therefore, we do not have the large dataset necessary to conduct deep learning unless we can obtain crowdsourced data sets from team projects in other PBL programs. As a proof-of-concept project, we use transfer learning to train the LA models with a small dataset. The basic idea is to use similar datasets (e.g., the Internet Movie Database of Maas et al., 2011), downloadable from Kaggle open datasets for sentimental analysis to pre-train the machine learning models. Then, we can use a small authentic, more relevant dataset such as the online discussion text data (see Chen, 2015; Spain et al., 2020) to fine-tune the model to achieve the desired accuracy. The similarity for pre-training data to the authentic data is evaluated by the input and output structures, such as the length of the bag of words for inputs and levels of sentimental analysis output. In contrast, the fine-tuning data have similar structures and similar keywords and topics of online course communication. Currently, the student co-authors used NPL and four RNN models to predict personality on social media. These models can be transferred to a LA application with a minor modification to support the automation of logged texts from BotCaptain. As a human-in-the-loop approach, our development advances parallel to the research effort presented in this paper and incrementally automates the data analytics capacity.

4.1 *Input data*

BotCaptain warehouses synchronous messages from teleconferencing meetings and asynchronous messages from social media into multiple databases through cloud services, as shown in Figure 3. The stored data structure is in JSON (Javascript Object Notation – a dictionary of key and value pairs) format, as shown in Figure 4. The NPL below groups all JSON messages under individual team members' names with the context of the team, weeks, conversation type, and time stamps.

Figure 3 The cloud services and databases for BotCaptain's data collection from social media

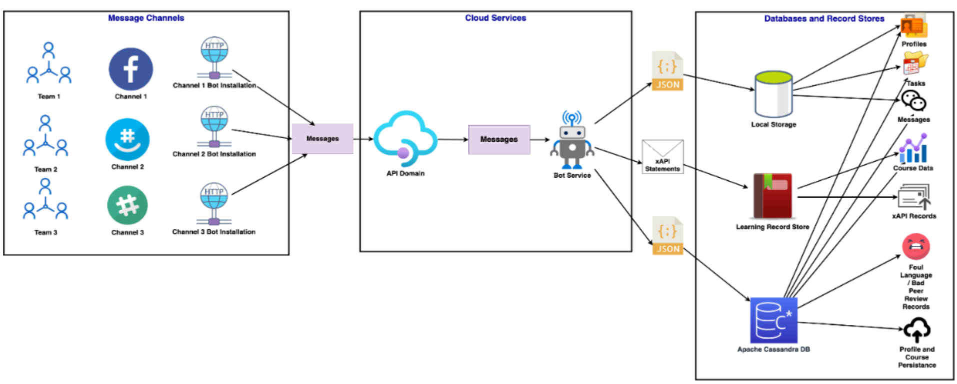


Figure 4 The JSON format of data


```
{
   "messages" [
        {     "time": "2020-06-03T14:36:58.178Z",
              "channelID": "emulator",
               "text": "t",
               "user": "User"
               "userID": "2e97949d - d527 - 48d3 - b989 -
5d67f",
               "team":  "team1",
                "nick": "Tim"
},
```


Note: The tags are channel ID, time stamps, user (articulator) information, and course and team identifications. The textual data's real content is short paragraphs of text in the place of the 't' field.

4.2 Natural language processing

The data frame structure depends on the task of data classification. We used Natural Language Toolkit (NLTK) Python library to pre-process data. Before using word embedding techniques to classify textual data, it is important to tokenise the words, remove punctuations, stop words, and stems, and finally transform the textual into standardised word vectors by following the Keras NLP Text Cleansing Procedure. Then, the content of textual message needs to be padded or cut into a sequence of texts of fixed length. As the labour-costly part of supervised learning, we need to manually label each data sample of our data frames for training and testing based on the classification tasks. In theory, if we can prepare labelled datasets that are large enough to train the models, all evaluations in Tables 1–8 can be automated. In practice, we need to use semi-supervised learning and transfer learning to prepare adequate laboured data for training models to the desired accuracy. This process has been just started. The two examples presented below use outside datasets to train our educational data mining (EDM) and LA models.

4.3 Naïve Bayesian model and support vector machine models for conversation classification

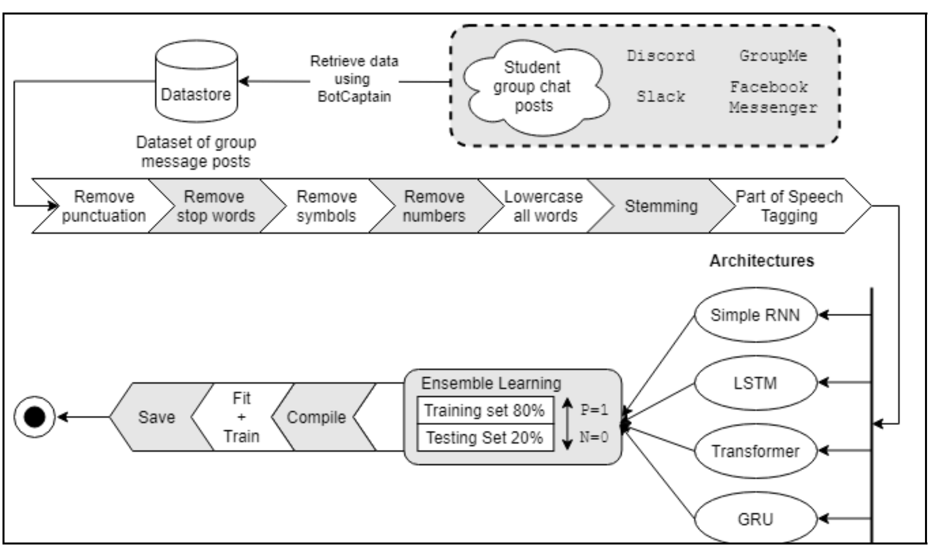
Table 8 for leadership and initiative depends on the messages that are classified as Trigger (T), Exploration (E), Integration (I), Resolution (R). The work described in this section reused the datasets by Chen et al. (2014), and the Java program one of our students wrote for the data mining course in the fall of 2015. The first author provided credits to the blog participation for homework, reading, and project assignments to motivate the students to participate in the blog-based discussions. However, the quantity of blogs reflects neither the students' true efforts of participation nor their understanding. To promote more quality conversation than just lengthy conversation in our online course at Syracuse University, Chen et al. (2014) manually identified the students' discussions and classified them into four phases: Triggers (T), Explorations (E), Integrations (I), and Resolutions (R), which reflect different contributions and depth of learning. She scored the full blogs of each student, both quantitatively and qualitatively. The Java program mentioned earlier used Chen et al.'s (2014) dataset of 773 blogs to classify the four stages of blog conversations. Once the textual data are transformed into a standard data frame, Naïve Bayesian and Support Vector Machine models provide the best performance to classify the four stages of conversations accurately. The four stages of task-oriented conversation for teamwork are defined in Chen, 2015 as Initiating, Diverging (brain-storm), Converging and Integrating (resolution). In the fall of 2017, another team of graduate students used the processed data and achieved 93% classification accuracy based on 10 folders of cross-validation method. These two previous works show that data mining can predict the stages of conversation reliably based on the frequencies of a bag of keywords relevant to the tasks and course contents. This method can be extended to classify other behaviour markers, such as the traits in Table 6 with small datasets. The shortcoming of this method is that it depends on manual labelling of all data and selecting keywords.

4.4 RNN models for sentimental textual analysis

Artificial neural network (ANN), recurrent neural network (RNN) in particular, has been successfully used for sentimental textual analysis in social media for decades. The same technology can be used to track if negative emotional statements that may harm the team spirit and social conflicts among teammates emerge; so that the instructor can intervene timely. To save time to label large datasets for the deep learning method, we use transfer learning and leverage the large open datasets to pre-train our RNN models.

In recent literature, supervised machine learning is used for personality trait identification (Asra and Shubhangi, 2015; Joshi et al., 2015). Bharadwaj et al. (2018) implemented a closely related study on personality traits based on the 16 types of Myers-Briggs Type Indicator (MBTI) personality questionnaire. The authors used a set of 8660 tweets to train a neural network to fit tweets. The vast MBTI9k dataset is aggregated from the MBTI Reddit pages known as subreddits by Gjurkovic and Snajder (2018). Non-neural network trials on MBTI9k prove that models using word n-gram features perform well with a large dataset. Supervised machine learning is proved as a valid approach for personality trait identification. Thus, we utilised recent Reddit user data for training a neural network and performing supervised learning for textual MBTI classification. As shown in Figure 5, four recurrent neural networks (RNN) architectures: Simple RNN, Gated Recurrent Unit, Long-Short-Term-Memory (LSTM), and LSTM with Attention. The LSTM model performed the best on the test set with a hamming distance score of 0.575 using the original dataset, while the LSTM with Attention model performed the best with a hamming distance score of 0.596 when using transfer learning. Our next step is to use the current models and pipeline as pre-trained models based on the Reddit dataset. We will then add another output layer as multitask RNN and use the available small data we collected to retrain the RNN models to evaluate the affective traits in Table 6 and interpersonal traits in Table 7. This is a typical transferring learning method to mitigate the problem of lacking adequate data to train a neural network.

Figure 5 Machine learning pipeline



4.5 Comprehensive evaluation based on machine output and peer evaluation

The outcomes of the LA described above only provide a series of snapshots of the behaviour markers associated with every individual and each team. We will combine the series of snapshots to make a comprehensive evaluation of every single individual and each team. A hybrid method will be used to prepare the data attributes listed in the rows in Tables 1–8. In this article, we only describe the individual evaluation, the evaluations of teams can be performed in a similar manner.

The four facets of traits are evaluated at three levels: A, B, and C. To evaluate a student's motivation, the instructor needs to fill out the first five items about accountability, efforts, grit, preparedness, and procrastination. The LA fills out the last item linked to promptness in replying to teammates. To prepare data for supervised learning when the training and testing datasets are large enough to achieve the desired accuracy, the instructor also needs to provide evaluations (A, B, or C) for each student at least over several semesters. The RNN models used in Table 6 to evaluate the affective trait are described above. The evaluation of interpersonal and social traits in Table 7 can be automated when the training dataset is large enough. However, the behavioural markers in Table 7 are more subjective than those in Table 6. The first four items of Table 8, for evaluation of leadership skills, can be automated by LA. The last item about the leadership role of technical tasks has to be filled in by the instructor or TA. Meeting minutes can be used as the ground truth to label the training datasets. In summary, all four traits of individual evaluations can be automated if we have the training dataset. The accuracy of the automated evaluation depends on the quantity and quality of the training datasets.

Once the four facets of traits for a student have been evaluated, the instructor obtains a holistic view of the student's overall contribution and effort to the team project. Then, the instructor can assign a fairly objective score based on the weighted total scores from Tables 5–8. The weighting for each item is at the instructor's discretion, depending on the nature of the project and course. The outcome of such a comprehensive evaluation is not limited to providing a fairer score for a team project, the associated xAPI record of LRS also provides a live reference letter for a student. The traits described in Tables 5–8 are the typical contents of the reference letters provided by instructors for students seeking internships or jobs. With the LRS, the instructor can save time writing truthful reference letters by simply sending a link to the LRS with the approval of the referees.

4.6 Outputs of the individual and team competence assessments in xAPI statement

This section illustrates a few examples of xAPI statements as the output of the individual and team competency assessments. Translating evaluation metadata into the xAPI format is a matter of utilising the correct verb, display, and definition. Because the exact words used in the four facets are not commonly used in the ADL directory, it is preferred to utilise a generic verb known as 'experienced' – along with a custom display and definition. This verb can be used for any situation as long as the display key is customised. The generic verb can be replaced by a semantically specific verb from a predefined domain-specific verb dictionary.

As shown in Figures 6–9, we illustrated the xAPI statements for tracking a student's accountability, grit, effort, and team meeting attendance. Because of xAPI's ubiquitous format statements can be cross-correlated with each other if using a well-developed LRS. As shown in Figure 10, translating form data into xAPI statements, however, can be a bit tricky at times. TinCanJS library is recommended for the task if the users consider a customised application. As the latest progress, Yet Analytics has developed a translating and simulation tool to automate such a task.

Figure 6 Example of tracking accountability


```
{} accountability_example.json > ...
{
    "actor":{
        "mbox":"bernart1@my.erau.edu",
        "Name":"Timothy Bernard"
    },
    "verb":{
        "id":"http://adlnet.gov/expapi/verbs/experienced",
        "display":{
            "en-US":"Accountability"
        }
    },
    "object":{
        "objectType":"StatementRef",
        "id":"8f87ccde-bb56-4c2e-ab83-44982ef22df0",
        "definition": {
            "name": { "en-US": "4" }
        }
    }
}
```


Figure 7 Example of tracking grit


```
{} grit_example.json > ...
{
    "actor":{
        "mbox":"bernart1@my.erau.edu",
        "Name":"Timothy Bernard"
    },
    "verb":{
        "id":"http://adlnet.gov/expapi/verbs/experienced",
        "display":{
            "en-US":"Grit"
        }
    },
    "object":{
        "objectType":"StatementRef",
        "id":"8f87ccde-bb56-4c2e-ab83-4498››2ef22df0",
        "definition": {
            "name": { "en-US": "4" }
        }
    }
}
```

Figure 8 Example of tracking efforts

efforts_example.json > ...


```json
{
    "actor":{
        "mbox":"bernart1@my.erau.edu",
        "Name":"Timothy Bernard"
    },
    "verb":{
        "id":"http://adlnet.gov/expapi/verbs/experienced",
        "display":{
            "en-US":"Efforts"
        }
    },
    "object":{
        "objectType":"StatementRef",
        "id":"8f87ccde-bb56-4c2e-ab83-44982ef22df0",
        "definition": {
            "name": { "en-US": "5" }
        }
    }
}
```


Figure 9 Example of tracking team meeting

halla22@my.erau.edu **attended** 'Team Meeting'


```json
{
    "id": "42ae532f-46c1-4c49-bd2b-2986c5800d46",
    "actor": {
        "objectType": "Agent",
        "mbox": "mailto:halla22@my.erau.edu"
    },
    "verb": {
        "id": "http://adlnet.gov/expapi/verbs/attended",
        "display": {
            "und": "attended"
        }
    },
    "timestamp": "2021-01-02T22:23:37.868Z",
    "stored": "2021-01-02T22:23:37.537Z",
    "authority": {
        "objectType": "Agent",
        "account": {
            "homePage": "http://cloud.scorm.com",
            "name": "UUVM0L7Q0T"
        }
    },
    "version": "1.0.0",
    "object": {
        "id": "http://icycle.org/botcaptain",
        "definition": {
            "name": {
                "en-US": "Team Meeting"
            },
            "description": {
                "en-US": "Meeting occured with a specific team in the class."
```

Figure 10 Example of TinCan.js

```
botcaptain > xAPI_Statements.js > xAPI_Statements > constructor
You, 2 minutes ago | 1 author (You)
var TinCan = require('tincanjs')
var jsonfile = require('jsonfile')
const config = jsonfile.readFileSync('./config.json')
const { bunlogger } = require('./bunlogger')
//Class for pushing statements into the LRS
You, 2 minutes ago | 1 author (You)
class xAPI_Statements {
    constructor(){
        try {
            this.lrs = new TinCan.LRS(
                {
                    endpoint: config.xAPI.endpoint,
                    username: config.xAPI.username,
                    password: config.xAPI.password,
                    allowFail: false
                }
            );
        } catch (ex) {
            console.log("Failed to setup LRS object: ", ex);
        }    You, 25 days ago • first commit
    }

    async recordLogin(email){
        if(config.xAPI.enabled == true){
            var statement = new TinCan.Statement(
                {
                    actor: {
                        objectType: "Agent",
                        mbox: `mailto:${email}`,
                    },
                    verb: {
                        id: "http://adlnet.gov/expapi/verbs/attended"
                    },
                    object: {
                        id: config.xAPI.objectID,
                        definition: {
                            name: {
                                "en-US": "Team Meeting"
                            },
```

5 Recommendations and metacognitive feedback

The LA using iCycle discussed above is still a pilot study, and more time is needed to tune the system for field testing. Besides its primary purpose to automate formative learning assessment for the non-cognitive team and individual competencies, the LA can also identify anomalies immediately. For example, once the sentimental textual analysis of LA identifies that the teammates exchange negative words passing a frequency threshold, the instructor will be noticed to investigate what is wrong and intervene timely. Also, it is expected that there will be a positive change in student behaviour change when students are aware of the data logs and their impact on their final grades. Empirical observation comparing CSCL teamwork with data logging and conventional teamwork without data logging indicates that the former significantly reduced teammates' conflicts, social loafing, and rude behaviours and languages (Oakley et al., 2004). This section will present examples of machined produced metacognitive feedbacks to students based on the individual competence assessments above. At the current stage, the instructor provides metacognitive feedback during the weekly meeting with team members. In future work, we will add features from iCycle to automate feedback as described below.

5.1 Metacognitive feedback based on individual competence assessments

The term metacognition is defined as the knowledge of one's own cognitive processes (Flavell, 1979). Over time, the definition of metacognition has evolved, and more comprehensive definitions have been proposed for the construct. More comprehensively, metacognition can be conceptualised as an awareness and knowledge of one's cognition and the ability to regulate it to ensure successful goal attainment (Schraw and Moshman, 1995). As reflected in the definition, metacognition has the following two key components: (1) an awareness of cognition and (2) the regulation of cognition. Awareness of cognition refers to the knowledge about cognitive tasks, strategies, and understanding learners possess about themselves and others (Flavell, 1979). The second component, regulation of cognition, refers to the process of regulating mental activities during learning through the following three distinct processes that play a critical role in the learning process: planning, monitoring, and evaluating. While planning involves all the strategies and time and effort expended before initiating a task, monitoring deals with understanding and performance during the learning process. Evaluating involves reflecting on one's performance after completing the task (Cross and Paris, 1988; Schraw and Moshman, 1995).

Given that metacognition positively influences the learning process, more researchers paid attention to investigate the role of metacognition in various learning environments recently. Consequently, more research studies emerge examining metacognition concerning the CSCL environments in which students are more likely to struggle to regulate their learning than in traditional learning environments (Volet et al., 2013). As expected, metacognition has been reported to foster successful collaboration in CSCL environments (Pifarré and Cobos, 2010; Volet et al., 2013). Moreover, providing metacognitive support and feedback to students in CSCL has been found to increase team productivity and interaction between team members (Kwon et al., 2013) and promote better cognitive and task performance (Pifarré et al., 2014).

The current project aims to foster successful teamwork in the CSCL environment by effectively monitoring individual team members' performance. Linking metacognitive support as feedback for CSCL can help improve the cumulative performance of every single member of the team (Ellis et al., 2013). In keeping with the individual competence model (see Tables 5–8), automated individualised metacognitive feedback, along with some form of formative feedback, will be delivered to team members based on one's effort and engagement throughout the team project. More specifically, to get automated individualised feedback delivered promptly to each team member, the data logged by BotCaptain and all other data sources, including progress reports, peer evaluation, and meeting minutes, will be consulted. Two individual dispositions toward teamwork, procrastination, and initiative, itemised in two different individual competence assessments, will be used to describe the nature of the two forms of automated feedback (metacognitive and formal) and their mode of delivery in more detail in the next few paragraphs.

One of the individual dispositions in the competence assessment is procrastination, which is itemised under the motivational aspect of the individual competence assessment. Procrastination is defined as the act of voluntarily delaying the performance and completing of an assigned task despite being aware of potential consequences (Steel, 2007). Numerous research studies have well documented the negative consequences of procrastination on performance in various domains (i.e., academic, social life, professional life). For example, procrastination in collaborative tasks has been found to result in poor academic functioning (Tice and Baumeister, 1997), missed deadlines (Ferrari, 1993), low quality of work (Grunschel et al., 2013), and increased levels of stress (Tice and Baumeister, 1997). In the current study, team communications, progress reports, and artefacts submitted in the shared team folder will serve as behavioural markers of procrastination. Team members who confirm or indicate completion or submission of an assigned task will be presented with the following automated formative prompt that serves as a reinforcement to keep them further engaged in the process: "Thank you for submitting your assigned task."

On the other hand, the BotCaptain will deliver metacognitive feedback to team members who exhibit one or two instances of delayed or incompleted submission to assigned tasks. Such feedback can help them plan effectively and openly monitor their progress (e.g., "You may think about breaking down your assigned task into manageable steps and be sure to start early to better allocate your time and effort"). Formative feedback in the form of a prompt such as "Please help your teammates stay on track by turning in your work promptly" is also necessary for team members who procrastinate. Based on the multiple data sources, a team member with more than two markers of procrastination (e.g., failing to upload an artefact or inquiring about possibly postponing the deadline) will receive automated metacognitive feedback in the form of hints. For example, a suggestion can be on how to monitor and eliminate distractions in the environment and focus on goals. In addition, a mechanised warning prompt featuring "the consequences of altering the work schedule/deadlines" is an example of formative feedback that will be automatically sent to individual team members who procrastinate more than once. Furthermore, the instructor will receive an automated message and be informed about a team member who exhibits three or more procrastination behaviours, as evidenced by the multiple data sources, so that the instructor can intervene to evaluate the case.

The initiative, on the other hand, as a leadership disposition, is listed as a trait under a team member's qualities of leadership and initiation (see Table 8). Initiative refers to an individual's active and self-motivated ability to work and go beyond what is formally required by a given task (Frese et al., 1996). Initiation has been found to be a predictor of creativity and is linked to idea generation (Binnewies et al., 2007), engagement, increased motivation, active performance, and perseverance in a task even if faced with obstacles (Frese and Fay, 2001). In our research project, the number of new ideas and discussion posts about a project, which is aggregated in the BotCaptain logs, will serve as the behavioural markers for a team member's initiative. To reinforce their behaviour and keep them engaged and motivated throughout the project, members who frequently share project ideas, make suggestions, and post comments will be presented with the following automated prompt: "Thank you for your contribution! Your leadership role in promoting your team is recognized." Team members who remain mostly silent and do not demonstrate any initiative by posting ideas, sharing thoughts, or starting a discussion for at least a week will receive metacognitive feedback in the form of prompts and questions. For instance, the following feedback sparks thinking and encourages the planning, monitoring, or evaluating cognitions. "What do you think your team should do next?" "Please share your thoughts", or "Please monitor your team's progress and then reflect on areas of improvement and how to improve the project." Besides, the students who exhibit less or no initiative will get formative feedback periodically as a checklist addressing their behaviours to increase metacognitive awareness about their insufficient involvement and remind them to be more active.

5.2 Data ownership, data security, and learner privacy issues

The persistent data saved in LRS contains private information similar to the information found in student transcripts. Therefore, data ownership, data security, and learner privacy issues need to be addressed publicly. As is the case with transcripts, the owners of the records in the LRS are each student and the academic institution that offers the PBL courses. Before the institution officially authorises its value, it may be delegate the institution's ownership temporarily to the course instructor. The institution or its designated delegators are responsible for protecting the students' privacy. This means that the personal learning record should not be released to any third party without the associated learner's permission or request. Data security is an issue and responsibility that the LRS vendors such as Watershed LRS, YetAnalytics, and Learning Locker addressed. Currently, the vendors such as Learning Locker provides free hosting of LRS for academic users. We will negotiate an affordable business model similar to those of other cloud services to provide unlimited and permanent services in the long term.

6 Conclusion and future work

With the increasing complexity of knowledge-based economics, teamwork is essential for future STEM workforces. Therefore, in postsecondary institutions, it is essential learning objectives to cultivate in students positive non-cognitive traits for successful teamwork. At Embry-Riddle Aeronautical University, PBL courses featuring team projects are used to develop students' non-cognitive competence and afford opportunities to collect data that make it possible to automate evaluations partially. This paper presents a task-agnostic team competence model, metadata, and the proof-of-concept technology iCycle for incrementally automating the evaluation of students' non-cognitive and task-agnostic skills in a CSCL environment with a focus on the teamwork process and educational technology that supports evaluation.

The CSCL environment provides opportunities for the instructors to collect a massive amount of data that provides a transparent view of the students' learning and performance. However, to our knowledge, there are very few open source data available in the EDM community to support the research presented in this paper, not to mention the large datasets needed to train and test the machine learning algorithms shown in Figure 5. The most serious challenge to the LA and EDM associated with this research is the quantity and quality of training datasets. At the current stage, the LA component is not ready to automate the data analysis described in this paper. A human-computer hybrid method is used to prepare such training datasets. This paper also presents a transfer learning method to mitigate the problem of lacking adequate authentic training data. In the near term, we see a heavy dependence on instructors to evaluate team processes. Nevertheless, over time, we expect greater automation as AI comes to play an increasingly critical role.

The success of EDU and LA in reducing instructors' workloads and providing students with more timely feedback depends on community-based research to share crowdsourced data in the EDM community. We cannot expect success from fullyautomated EDM and LA in the short term. This research serves as a step to synergising human effort and AI-based educational technology to reduce instructors' workload and provide students with more timely feedback on their performance in a CSCL environment. As for future work, as shown in the shaded yellow box of Figure 1, we plan to customise a third-party virtual TA called LearnPal with a Q&A feature. The LearnPal may answer simple task-agnostic questions but forward the task-specific and conceptual questions to instructors or TA. We will use the interoperable data exchange technology xAPI for BotCaptain to exchange data with LearnPal and the intelligent tutoring tools GIFT (Liu et al., 2020). Other future work will integrate iCycle with the Competency and Skills System (https://cass.extension.eduworks.com/) of a Total Learning Architecture (https://adlnet.gov/projects/tla/). Once the integration is proven successful, we will be able to replace most of the in-housing developed EDU and LA technology.

References

Asra, S. and Shubhangi, D.C. (2015) 'Personality trait identification using unconstrained cursive and mood invariant handwritten text', *International Journal of Education and Management Engineering*, Vol. 5, No. 5, p.20.

Bharadwaj, S., Sridhar, S., Choudhary, R. and Srinath, R. (2018) 'Persona traits identification based on myers-briggs type indicator(mbti) – a text classification approach', *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp.1076–1082.

Biasutti, M. and Frate, S. (2018) 'Group metacognition in online collaborative learning: validity and reliability of the group metacognition scale', *Education Technology Research Development*, Vol. 66, pp.1321–1338. https://doi.org/10.1007/s11423-018-9583-0

Binnewies, C., Ohly, S. and Sonnentag, S. (2007) 'Taking personal initiative and communicating about ideas: What is important for the creative process and for idea creativity?' *European Journal of Work and Organizational Psychology*, Vol. 16, No. 4, pp.432–455.

Brenna, F., Danesi, G., Finch, G., Goehring, B. and Goyal, M. (2018) *Shifting toward enterprise-grade AI: resolving data and skills gaps to realize value*, IBM Institute for Business Value. Available online at: https://www.ibm.com/downloads/cas/QQ5KZLEL

Burmester, E. (2020) 'Authoring collective training demonstrations in GIFT', *Proceedings of the 8th Annual GIFT Users Symposium (GIFTSym8).*

Chen, L., Gonzalez, E. and Nantermoz, C. (2014) *Authorship attribution with limited text on twitter*. Available online at: http://cs229.stanford.edu/proj2017/final-reports/5241953.pdf

Chen, Y. (2015) 'Automated content analysis of students' cognitive presence in asynchronous online discussion', *Proceedings of World Conference on Educational Media and Technology 2015*, Association for the Advancement of Computing in Education (AACE), pp.339–344.

Cross, D.R. and Paris, S.G. (1988) 'Developmental and instructional analyses of children's metacognition and reading comprehension', *Journal of Educational Psychology*, Vol. 80, No. 2, pp.131–142.

Ellis, A.P.J., Hollenbeck, J.R., Ilgen, D.R., Porter, C.O.L.H., West, B.J. and Moon, H. (2003) 'Team learning: collectively connecting the dots', *Journal of Applied Psychology*, Vol. 88, pp.821–835.

Ferrari, J.R. (1993) 'Christmas and procrastination: explaining lack of diligence at a "real-world" task deadline', *Personality and Individual Differences*, Vol. 14, pp.25–33.

Flavell, J.H. (1979) 'Metacognition and cognitive monitoring: a new area of cognitive developmental inquiry', *American Psychologist*, Vol. 34, pp.906–911.

Frese, M. and Fay, D. (2001) 'Personal Initiative (PI): an active performance concept for work in the 21st century', in Staw, B.M. and Sutton, R.M. (Eds): *Research in Organizational Behavior*, Vol. 23, pp.133–187.

Frese, M., Kring, W., Soose, A. and Zempel, J. (1996) 'Personal initiative at work: differences between East and West Germany', *Academy of Management Journal*, Vol. 39, No. 1, pp.37–63.

Gjurkovic, M. and Snajder, J. (2018) 'Reddit: a gold mine for personality prediction', *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pp.87–97.

Grunschel, C., Patrzek, J. and Fries, S. (2013) 'Exploring the reasons and consequences of academic procrastination: an interview study', *European Journal of Psychology of Education*, Vol. 28, pp.841–861.

Joshi, P.M., Agarwal, A., Dhavale, A., Suryavanshi, R. and Kodolikar, S. (2015) 'Handwriting analysis for detection of personality traits using machine learning approach', *International Journal of Computer Applications*, Vol. 130, pp.40–45.

Kwon, K., Hong, R.-Y. and Laffey, J.M. (2013) 'The educational impact of metacognitive group coordination in computer-supported collaborative learning', *Computers in Human Behavior*, Vol. 29, No. 4, pp.1271–1281.

Liu, H., Bernard, T. and Acharya, K. (2020) 'Using GIFT to develop adaptive remedial courses for graduate degree programs in data science', *Proceedings of the 8th Annual GIFT Users Symposium*, pp.61–68.

Liu, H., Spector, J.M. and Ikle, M. (2018) 'Computer technologies for model-based collaborative learning: a research-based approach with initial findings', *Computer Application for Engineering Education*, pp.1–10. https://doi.org/10.1002/cae.22049.

Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A. and Potts, C. (2011) 'Learning word vectors for sentiment analysis', *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, pp.142–150. Available online at: http://www.aclweb.org/anthology/P11-1015

Mathieu, J., Maynard, M.T., Rapp, T. and Gilson, L. (2008) 'Team effectiveness 1997-2007: a review of recent advancements and a glimpse into the future', *Journal of Management*, Vol. 34, No. 3, pp.410–476.

McCormack, R., Case, A., Howard, D., Logue, J., Kay, K. and Sinatra, A.M. (2020) 'Teamwork training in GIFT: updates on measurement and audio analysis', *Proceedings of the 8th Annual GIFT Users Symposium (GIFTSym8)*.

Messick, S. (1994) *The interplay of evidence and consequences in the validation of performance assessments*, Educational Testing Service, Princeton, ETS-RR-92-39. Available online at: https://files.eric.ed.gov/fulltext/ED390891.pdf

Oakley, B., Felder, R.M., Brent, R. and Elhajj, I. (2004) 'Turning student groups into effective teams', *Journal of Student Centered Learning*, pp.10–34.

Olivares, O.J. (2008) 'Collaborative vs. cooperative learning: the instructor's role in computer supported collaborative learning', in Orvis, K.L. and Lassiter, A.L.R. (Eds): *Computer-Supported Collaborative Learning: Best Practices and Principles for Instructors*, pp.20–38. Available online at: https://www.researchgate.net/publication/200027060

Orvis, K.L. and Lassiter, A.L.R. (Eds) (2008) *Computer-Supported Collaborative Learning: Best Practices and Principles for Instructors*, Information Science Publishing. Available online at: https://www.researchgate.net/publication/200027060

Pifarré, M. and Cobos, R. (2010) 'Promoting metacognitive skills through peer scaffolding in a CSCL environment', *Computer Supported Learning, Technology and Learning*, Vol. 5, pp.237–253.

Pifarré, M., Cobos, R. and Argelagós, E. (2014) 'Incidence of group awareness information on students' collaborative learning processes', *Journal of Computer Assisted Learning*, Vol. 30, No. 4, pp.300–317.

Rosé, C.P., Howley, I., Wen, M., Yang, D. and Ferschke, O. (2017) 'Assessment of discussion in learning contexts', in von Davier, A., Zhu, M. and Kyllonon, P. (Eds): *Innovative Assessment of Collaboration*, Springer, New York, NY, pp.81–94.

Schraw, G. and Moshman, D. (1995) 'Metacognitive theories', *Educational Psychology Review*, Vol. 7, No. 4, pp.351–371.

Smith-Jentsch, K.A., Cannon-Bowers, J.A., Tannenbaum, S.I. and Salas, E. (2008) 'Guided team self-correction: impacts on team mental models, processes, and effectiveness', *Small Group Research*, Vol. 39, No. 3, pp.303–327.

Sottilare, R., Burke, S., Sinatra, A.M., Johnston, J. and Salas, E. (2020) *Toward a Scientifically Rooted Design Architecture of Team Process and Performance Modeling in Adaptive, Team-Based Intelligent Tutoring Systems*, Technical Report.

Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H. and Gilbert, S.B. (2018) 'Designing adaptive instruction for teams: a meta-analysis', International Journal of Artificial Intelligence in Education, Vol. 28, No. 2, pp.225–264.

Spain, R., Min, W., Saville, J., Mott, B., Brawner, K., Johnston, J., Goodwin, G. and Lester, J. (2020) 'Team communication analytics using automated speech recognition', *Proceedings of the 8th Annual GIFT Users Symposium (GIFTSym8)*.

Spector, J.M. and Wu, J. (2017) 'Artificial intelligence and the future of education: big promises - bigger challenges', *US-China Smart Education Conference*, Beijing Normal University, March.

Steel, P. (2007) 'The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure', *Psychological Bulletin*, Vol. 133, pp.65–94.

Tice, D.M. and Baumeister, R.F. (1997) 'Longitudinal study of procrastination, performance, stress, and health: the costs and benefits of dawdling', *Psychological Science*, Vol. 8, pp.454–458.

Volet, S., Vauras, M., Khosa, D. and Iiskala, T. (2013) 'Metacognitive regulation in collaborative learning: conceptual developments and methodological contextualizations', in Volet, S. and Vauras, M. (Eds): *Interpersonal Regulation of Learning and Motivation: Methodological Advances*, Routledge, New York, NY, pp.67–101.