Employability Skills: Closing the Gap Between Academia and Industry Through Personalized Student Journey Maps

Group 9

Jiayao Tong (Student Id: 958905)

Simin Zuo (Student Id: 1199404)

Vanshree Bapat (Student Id: 1208561)

Wei Wang (Student Id: 900889)

Yixuan Liu (Student Id: 980551)

1 Introduction

Education when applied in form of employment (self or other) helps an individual pave their way in life. Philosophically, employment is not the only way to lead a comfortable life, but in practical sense, some method of income generation is of utmost importance. Having the skills to apply the education you receive is the ideal goal. Students often find themselves hunting for jobs after the completion of their education. There are many confounding factors affecting this but one of the major factors is, not being able to crack the interviews. Not being able to practically apply the knowledge gained is one serious concern while applying for jobs. Even from the educational institute's perspective, the graduate employment rate is often used to assess the quality of education provided.

The employability rate can be improved by tailoring the curriculum of the educational institute as well as providing feedback to the students on their performance and guiding them through the coursework. Our project focuses on the student aspect of this equation. The core of the project is applying Self-regulated learning concepts to help measure the employability of a student and to create an individual journey map of their progress, to help them understand the areas they need to focus on as well as provide a quantifiable encouragement on their work.

It is always beneficial for the educational institute to appropriately assess the learning outcomes of its curriculum and keep it updated as per the current employability market. Helping students by providing a clear journey map of their progress and acquisition of employability skills will definitely be beneficial for the students as well as the institute. This project helps in these two areas using the data extracted from Canvas Learning Management System. This data is only extracted from student who provide consent to participate in this study. It collects data about the assignments, modules, grades and most importantly the number of times a page has been viewed (page views).

The project has been partially developed with the SRL measurement matrix created and implemented. A dashboard reflecting the students' progress and behaviours with respect to the LMS has also been launched. Our part now includes analysing the data, finding patterns, and generating prediction to compare with actual results.

The final aim of this project is to use the Canvas data to develop an analytic learning tool to enhance students' outcomes and employability, create progress and evidence maps as part of student's journey map, which supports the visualisation skills at the subject level and award credits that could be used by students and industry alike using scaled assessments.

2 Related Work

SRL (self-regulated learning) is recognised as a vital part of ICT (Information Communication Technology) competence and is positively correlated (Khampirat, 2021). In this era of technological explosion, hirers expect candidates to have sufficient ICT competence to handle the challenges at the workplace (Abbasi et al., 2018; Panadero & Alonso-Tapia, 2014; Sarkar et al., 2020). Furthermore, students and educational institutions have begun to be aware that SRL is the basis for achieving academic results and avoiding academic failure in higher education (Hsu et al., 2021; Lobos et al., 2021; Poon, 2014). In addition, self-regulated learning also has a strong correlation with employment, especially early career maturity. In other words, self-regulated learning plays a role of mediator between career maturity and perceived employability (Hsu et al., 2021).

The core of SRL is the application of self-regulatory strategies to keep one's behaviour on task and psychologically motivated according to a predetermined plan (Panadero & Alonso-Tapia, 2014). Although many SRL models are available, they divide the entire SRL process into several distinct stages and are cyclical (Kitsantas & Zimmerman, 2008; PUUSTINEN, 2001). In Zimmerman's model, SRL is divided into three phases: forethought, performance, and self-reflection (Panadero & Alonso-Tapia, 2014). In other words, SRL can be improved step by step during a learning cycle and enhanced in each cycle. Therefore, it is a reasonable research method to track whether the SRL and academic performance of the same students are optimally improved over a more extended period and several learning cycles. In Xiao and Yang's research (2019), experts from the research station came to the school twice in two consecutive years to collect continuous data from the same participants to study the relationship between SRL and formative assessment, which means that they thought it would be more meaningful to study the changes in SRL skills in the same students. SRL skills are an abstract concept and can only be reflected in changes and improvements in SRL skills through changes in participants' behaviours over time. Some studies have used tests as the assessment time point (pre-test and post-test) to compare changes in participants' SRL skills (Lobos et al., 2021; Susaki, 2021), in addition to correlation tests between examination results and SRL skill performance behaviours (Cosnefroy et al., 2018).

It is also reasonable to divide the participants into intervention and control groups. In Susaki's (2021) research, students' scores in dribbling skills and the self-regulation strategies were recorded at the time point of pre-test and post-test to compare whether receiving SRL aid in the learning process resulted in student progress and whether the performance of students in the intervention group improved significantly compared to those without SRL aid. However, the results showed no significant difference between the scores of the control and intervention groups. This research does not consider individual variability in that each person's SRL abilities vary. The methodology of Lobos's (2021) study seems more rigorous, while the participants were segmented into experiment and control groups as well. However, researchers set variance homogeneity and sphericity assumptions to ensure that the two test

groups have similar covariance matrices and the same variance differences, which could prevent the drawbacks of Susaki's (2021) work effectively.

There are several methods regarding data acquisition and data collection, such as questionnaires, interviews, observations, clickstream data. It has been argued that questionnaires are the best tool to measure and assess SRL (Maldonado-Mahauad et al., 2020), and this is true, as many studies on SRL have used questionnaires (Cosnefroy et al., 2018; Maldonado-Mahauad et al., 2020). However, the limitation of questionnaires is obvious, the choices that participants choose may be subconscious and not conform to the reality, which may cause the statistical errors and deviation. Since the impact of the epidemic and the popularity of online courses, many students attend classes and study via the Internet. That is to say, log data of their access to the web and performance data on the study site can be obtained somehow (Qiao et al., 2021). Compared with the participants' behaviours that reflect by themselves, these data can be more comprehensive and continuous for deeper research, and these statistic data can be better visualised than the questionnaire results (Lim et al., 2020). Moreover, the data used in this study should be variable, otherwise may cause the problem of potential self-report bias. Hence, for completing deeper research, collecting data from different sources is necessary.

It is widespread in research to perform simple fundamental statistical analysis of the collected data, such as correlation and descriptive statistics, before going into more profound questions and adopting more complex methods (Cosnefroy et al., 2018; Hsu et al., 2021; Mih & Mih, 2011). As mentioned above, SRL is an abstract and unmeasurable concept. Some researchers use structural equation modelling to model the path between SRL and other factors (Hsu et al., 2021; Lim et al., 2020). It combines factor analysis, variance analysis and multiple regression (Klein, 2016) to explain the structural relationship between latent variables as well as the relationship between measured variables and latent variables (Korstanje, 2021). Other studies use EFA (Exploratory Factor Analysis) to measure and explain SRL and the factors they focus on at the interval level (Cosnefroy et al., 2018; Maldonado-Mahauad et al., 2020). It is based on correlation matrix between the factors and is more suitable for studying multidimensional problems like SRL (Fontaine, 2004). In addition to these experimental design surprises, generative models like the Hidden Markov Model can also be used to analyse the data. In Qiao's research (2021), the participants were divided into three categories according to the collected data, and Hidden Markov Models were established for the behaviour data of the three types of SRL learners to study the behaviour sequence and habits of these three types of learners at various stages of the SRL process.

3 Data Analysis

3.1. Overview of the project

This project is based on the contribution of Self-regulated Learning in the employability of an individual. The core of the project is made of 3 activities:

1) Monitoring the progress of a student is done using a matrix which has been provided to us by the clients and was used in our preliminary analysis.
2) Analysing behavioural patterns in the collected student data which will be performed by our team in the next semester.
3) And generating suggestions for improvement of a student, will be based on the results of our models and after consultation with the client.

These 3 points act as a guide for the academic and employability betterment of an individual.

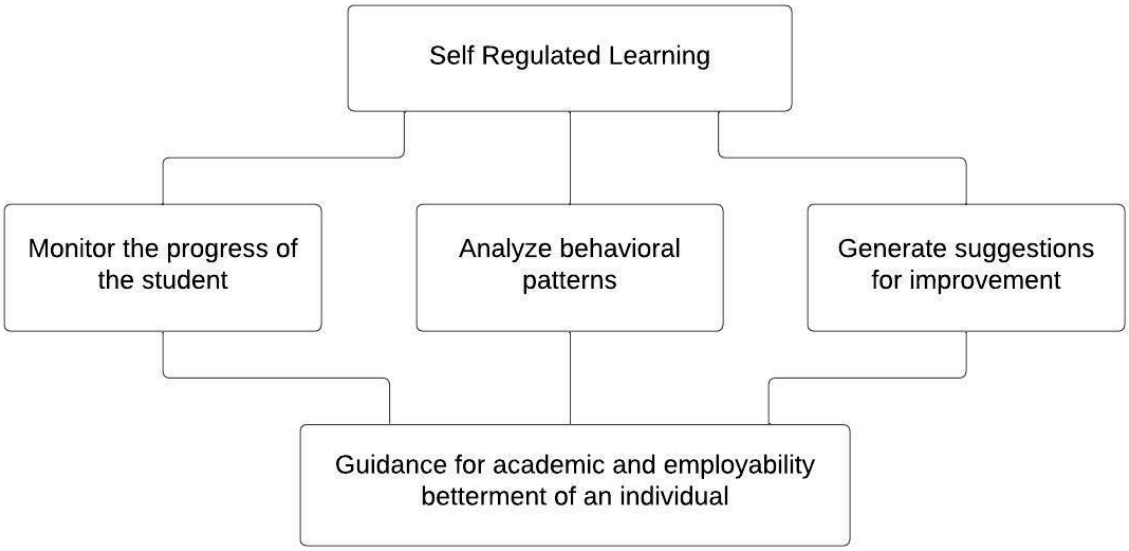


Figure 1 Project conceptual structure

In agreement with the client, we shortlisted three subjects as the basis for our project. Our next step was to extract the list of students enrolled in these subjects and find students that were common either in all three subjects or in combination of two subjects. We then proceeded to extract data for these students as per the requirements of the marking functions. These marking functions have been provided to us by the client and are explained in detail in this report. The analysis done for the first item has been explained as well. Similar analysis can be extended to the other items as well.

3.2. Marking Functions

The following items have been designed by the client and act as a score matrix to measure the SRL impact. The criteria column in the below matrices holds the descriptions of the score achieved, and the score contains the score as per the conditions of the matrices.

Item 1 – Examines Canvas at the outset of the semester to prepare for learning

Variables:

1 Major assessments: Number of assessments are in each course.

2 Lecture Slides and Readings: Number of Lecture slides and readings in each course.

3 Discussion board and People tab: If the student checked how to connect with others.

Criteria	Score	How to achieve the criteria (within week 0-3 of the semester) (Threshold Percentage: x = 100%)		
		Views x% of assessments	Views x% of Modules	Views x% of Discussion Forum posts and/or People Page

Figure 2 Score calculation formula for item 1

This item measures the activity of students in the first 3 weeks of the semester. It measures if the students are taking efforts to explore the subject on LMS and keep up with the content taught. It considers three variables namely, assignment page views, Content (lectures and reading material) views and discussion forum views. The score ranges between 0 to 3, and 1 score unit is generated only if all pages of the said variable have been visited i.e., 100% of the accessible pages have been accessed.

Item 2 - Paces learning across the semester

Variables:

1 Day Type: Indicates the expected behaviour in
Normal - Day is not within 3 days of assessment
Lead Up - Day is within 3 days of assessment

2 Sessions (Long / Short): Indicates the amount of time spent browsing canvas.
Long Session: Greater than or equal to 50 mins
Short Session: More than 30 mins

Criteria	Score	Condition			
		Day Type == Normal (Normal periods are grouped in sets of 7 days)		Day Type == Lead Up (Assessment lead up periods are grouped independently)	
		Number of short sessions >= 1 for every 7 normal days	Number of long sessions >= 1 for every 7 normal days	Number of short sessions >= 1 for every assessment leadup up period	Number of long sessions >= 1 for every assessment leadup up period

Figure 3 Score calculation formula for item 2

This item measures the study pattern of the students. It considers the amount of time spent on the LMS, mainly while viewing lectures. The score is measured depending on the number of long (≥50 mins) or short (>30 mins) sessions during the normal days (days other than the 3 days prior to assignment submission) and the lead up days (3 days prior to assignment submissions). This item will act as a criterion in trend detection in the study patterns of the class, which is one of our action items in the next semester.

Item 3 - Reads instructions of assessments

Variables:

1 Number of Assessments
2 Number of views for each Assessment
3 First viewing threshold: This variable is a time-period that takes place immediately after the assessment instructions have been released. If a student views the instructions in this time-period, it could mean they are planning ahead.

Criteria	Score	How to achieve the criteria	
		All assessment instructions are viewed before completion >= 1 times	All assessment instructions are viewed before completion >= 2 times at least 24 hours apart

Figure 4 Score calculation formula for item 3

This item helps understand how a student plans for an assignment. It measures scores based on the number of views within a particular time frame. Using this, we might be able to analyse student's planning, and separate them into different classes based on grades, to see if students belonging to a particular class display similar planning patterns.

Item 4 - Completes extra/formative assessment in the course

Variables:

1 Extra Assessment: An assignment that does not contribute to the student's final grade.
2 Views X% of extra assessment before completion: The percentage of extra assessment viewed before its completion date exceeds some threshold X.
3 Completes X% of extra assessment: The percentage of extra assessments completed exceeds some threshold X.

4 Views X% of extra assessment after completion: The percentage of extra assessment viewed after its completion date exceeds some threshold X.

Criteria	Score	Condition		
		Views at least one extra assessment before completion	Completes at least one extra assessment	Views at least one extra assessment after completion

Figure 5 Score calculation formula for item 4

This item helps understand how a student treats the extra assessment that does not contribute to the student's final grade by checking the number of views during the assessment, number of completion and number of views after the completion for this extra assessment. To see if there is any relevance between the scores students received for other assignments and students' behaviour for this extra assessment.

Item 5 - Pulls together information from various sources when studying

Variables:

1 The average number of sources referenced per study session: The number of readings and discussions accessed on average during every defined study session of duration X.
2 Source Type: Includes Discussion, Module, Quiz and Lecture Capture

Criteria	Score	Condition	
		Average number of source types referenced per study session	
		>= 1	>= 2

Figure 6 Score calculation formula for item 5

This item is developed to understand how students use external materials while doing their assignments by checking the average number of source types referenced per study session. To see the relationship between the number of extra external materials students used and the final score they have received for this assignment.

Item 6 - Reviews materials across the semester to understand important relationships between them

Variables:

1 Whether or not the student has viewed the specified percentage of lectures and readings
2 Day Type: Indicates the expected behaviour in
Normal - Day is not within X days of assessment
Lead Up - Day is within X days of assessment

3 Views at least one new lecture or reading in the same session as a prior lecture or reading

Criteria	Score	Condition		
		Accesses at least at least one file TODO (in future could change this to 80% of all files)	Mainly access files on normal days (not lead up days)	Views at least 20% of FILES in the same SESSION as another (prior) FILE Sessions where student reviews past material / total num sessions where the student views a new file >= 20%

Figure 7 Score calculation formula for item 6

This item helps to understand how students review materials by checking the percentage of lectures and readings related to this assessment students have viewed. We can see the relationship between the scores students received and the number of reviews for relevant materials using this.

Item 7 - Follows a weekly study schedule

Variables:

1 The average number of study sessions per week
2 If student accesses canvas every week
3 If a student follows a weekly schedule
4 The final score for this subject

Criteria	Score	How to achieve the criteria		
		Average number of study sessions per week >= 3	Accesses canvas every week	Follows weekly schedule

Figure 8 Score calculation formula for item 7

This item helps understand how students perform by following the schedule by checking the average number of study sessions per week, if the student accesses canvas every week and if the student follows a weekly schedule—using this to see the relationship between students' final score for this subject and their scheduling abilities.

3.3. Description of Data

Before describing the data, it is worth introducing the method to extract it. At the beginning, the client recommended we use Canvas API to retrieve log data from the Canvas database through the Postman platform. The API could retrieve all data about the information on courses and students and all students' behaviours on Canvas. In this way, following the items in the marking function, we utilized the request package in python to query all the data we needed. One thing that needs to be mentioned is that the extraction cannot be done by simply requesting the URL. Only 100 rows of data could be retrieved for each URL. To get the URL for the next page, we need to retrieve it from the head of the last URL and get all data iteratively. The figure below provides the solution we found to solve the problem above, extracting records as much as we want.

```
## pseudo code
# get page view of a student by its id from the initial url (first page)
url = f"https://unimelb-prod.instructure.com/api/v1/users/{id}/page_views?per_page=100"
r = requests.get(url, headers=headers)
result = response.json()

# get the url for next page
linkheader = r.headers['Link']
# result example
"""
<https://unimelb-prod.instructure.com/api/v1/users/[redacted]/page_views?page=bookmark:WyIyMDIyLTAzLTEzVDE0OjUx
OjM3LjA4MCsxMTowMCIsIjEwZmJmZGM5LTVkNzYtNDk2ZS04YzczLWUzMzgzYzA5MTYyNiJd&per_page=100>; rel="current",<ht
tps://unimelb-prod.instructure.com/api/v1/users/[redacted]/page_views?page=bookmark:WyIyMDIyLTAzLTA2VDExOjE4OjA
zLjMzMCsxMTowMCIsImJkYTI2ZjIzLTliNTMtNDE0MC1hMjRmLTcyNDM5YTRkYzliYyJd&per_page=100>; rel="next",<https://
unimelb-prod.instructure.com/api/v1/users/[redacted]/page_views?page=first&per_page=100>; rel="first"
"""
# use the next url to find next page and continue until the last page
```

Figure 9 Pseudo code for retrieving page view of a student

After retrieving the data, the first thing we should consider is finding the information of each course and the students' information for that course. The super token given by the client enabled us to access the data of 3 subjects during three consequential semesters, which are all compulsory subjects in the course structure of the Master of Software Engineering. In other words, it is possible to find common students in two or all the three subjects. Figure 2 indicates the student distribution among the three subjects. Then, the subsets were generated from the students' lists of the three subjects, denoted as subjects A, B and C, illustrated below in Table1, including the size of each subset. After extracting the user id of the selected population, we utilized the user id and course id to retrieve the grades of specific students in specific subjects through the URL under the Enrolment folder.

Subsets	Size
A∩B	45
A∩C	44
B∩C	62
A∩B∩C	42

Table 1 subsets of population and the size of each subset

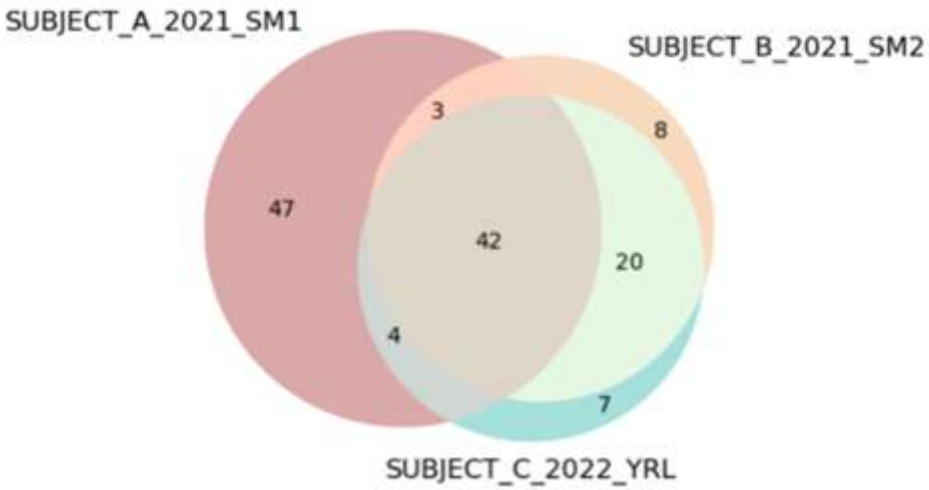


Figure 10 Average pageviews for 3 activities

In addition, we use the user id to find the pageview records of each student, which is a type of log file. It is worth mentioning that the data we get through the API is in JSON format. First, we converted the JSON file to a CSV file. The pageview raw data contains 22 columns, and the 'URL' and the time columns are used to indicate what students did on Canvas at a particular time, which means it is possible to monitor students' learning behaviour on Canvas. Afterwards, we extract information from the URLs according to the marking function given by the client. What learning behaviours URLs represent are listed in Table 2.

	Id	Activity	URL
user	1	check account notification	https://canvas.lms.unimelb.edu.au/api/v1/accounts/self/users/self/account_notifications?per_page=100
	2	check course announcement	https://canvas.lms.unimelb.edu.au/courses/{couse_id}/announcements
	3	log in	https://canvas.lms.unimelb.edu.au/api/v1/users/self/enrollments
course	4	check course	https://canvas.lms.unimelb.edu.au/courses/{course_id}
	5	check the course modules	https://canvas.lms.unimelb.edu.au/courses/{course_id}/modules
course	6	check a specific content in the modules	https://canvas.lms.unimelb.edu.au/api/v1/courses/{course_id}/module_item_sequence?asset_type=ModuleItem&asset_id={content_id}
	7	check discussion topic	https://canvas.lms.unimelb.edu.au/courses/{course_id}/discussion_topics
	8	check file previews	https://canvas.lms.unimelb.edu.au/courses/{course_id}/files/{file_id}/file_preview
	9	download files	https://canvas.lms.unimelb.edu.au/courses/{course_id}/files/{file_id}/download
	10	check folders	https://canvas.lms.unimelb.edu.au/api/v1/users/self/folders/root

	11	check grades	https://canvas.lms.unimelb.edu.au/courses/{course_id}/grades
Assignment	12	check assignment	https://canvas.lms.unimelb.edu.au/courses/{course_id}/assignments
	13	check the content of an assignment	https://canvas.lms.unimelb.edu.au/api/v1/courses/{couse_id}/assignments/{assignment_id}
	14	download assignment description	https://canvas.lms.unimelb.edu.au/courses/{course_id}/assignments/{assignment_id}/submissions/120306?download={download_id}
	15	submit assignment	https://canvas.lms.unimelb.edu.au/courses/{course_id}/assignments/{asasignment_id}/submissions
Quiz	16	check quiz	https://canvas.lms.unimelb.edu.au/api/v1/courses/{course_id}/all_quizzes?per_page=100
	17	check the content of a quiz	https://canvas.lms.unimelb.edu.au/api/v1/courses/{course_id}/quizzes/{quiz_id}
	18	do quiz	https://canvas.lms.unimelb.edu.au/api/v1/courses/{couse_id}/quizzes/{quiz_id}/submissions?per_page=100

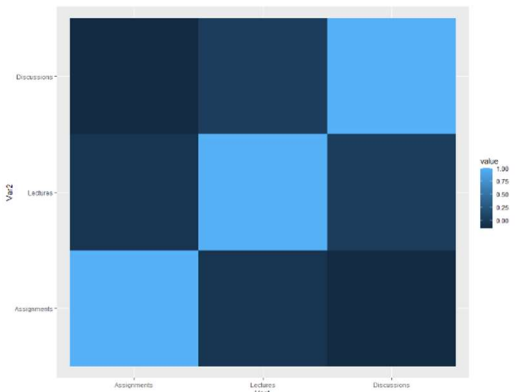
Table 2 URLs and the categories of each activity

After locating information, we need to divide different activities into several categories referring the marking function, and then count the activities under each category. Below is an example of how we can get data from pageview records according to the 1st item in marking functions.

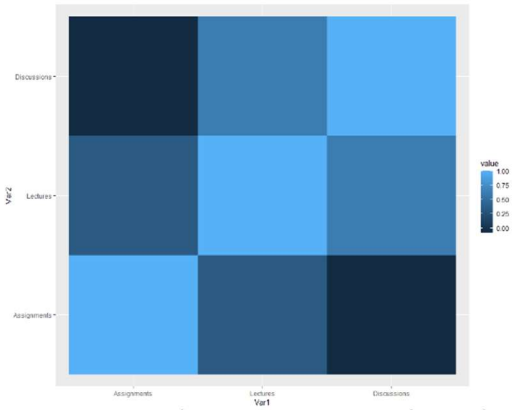
For 1st item, we need to extract information from id 6 (for lectures and readings), 7 (for discussion board) and 14 (for assignment). As the item requires, we only focus on the pace of students from week 0 to week 3 for each semester. For id 6 and 14, we need to extract the lecture/reading/assignment ids from the URL (student's behaviours) and compare them with all lecture/reading/assignment ids we extract from the course information URL (course information). Therefore, we can find out how many lectures/readings/assignments have already been checked and when did they checked. For id 7, we limit the time duration from week0 to week3 and count how many times the URL is visited. If the student visits the particular URL enough times to meet the criteria, he will get the corresponding score. Similar mechanisms will be implemented for other items.

3.4. Data Analysis Example

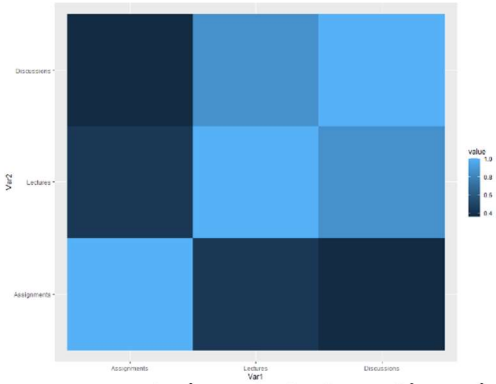
According to the data extract from the URL, following analysis has been reported. It is known that the marking function were used to generate score depending on student's behaviour. Therefore, the preliminary analysis is based on the correlation between each variable of a particular item. Here is the example of data analysis we have done, which is focuses on item1. As shown in the last section, there are 3 variables in item1, page views of assessment, course materials and discussion board for each student in 3 subjects.



	Assignments	Lectures	Discussions
Assignments	1.00	-0.06	-0.16
Lectures	-0.06	1.00	0.02
Discussions	-0.16	0.02	1.00



	Assignments	Lectures	Discussions
Assignments	1.00	0.33	-0.08
Lectures	0.33	1.00	0.61
Discussions	-0.08	0.61	1.00



	Assignments	Lectures	Discussions
Assignments	1.00	0.43	0.36
Lectures	0.43	1.00	0.86
Discussions	0.36	0.86	1.00

Figure 11 Heatmaps for Subject A,B,C

Initially to analyze the correlation of different variables in every subject, we prefer to use heatmap to reflect and visualize. The figures above show the results for these 3 subjects.

In subject A, the correlations between variables display unrelated, even some variables such as assignment versus lecture and assignment versus discussions are inversely proportional. However, the correlations between variables for subject B and subject C is positive, especially in subject C, there is a strong correlation between lecture versus discussion and lecture versus assignment. Compared the heatmap of subject B with subject C, the correlations between variables are similar mostly, although there is a little difference of degree of correlation, we can observe that there are shades of colours. Therefore,

it can be concluded that these 3 variables have the stable and positive correlation, which proves that it is reasonable in marking function to gather them in one item to reflect the ability of students.

As mentioned in the marking function, the scoring is based on the views of each variable. Therefore, we did the statistic about the general trend of students accessing canvas in each subject, and the analysis result is reflected through bar plot.

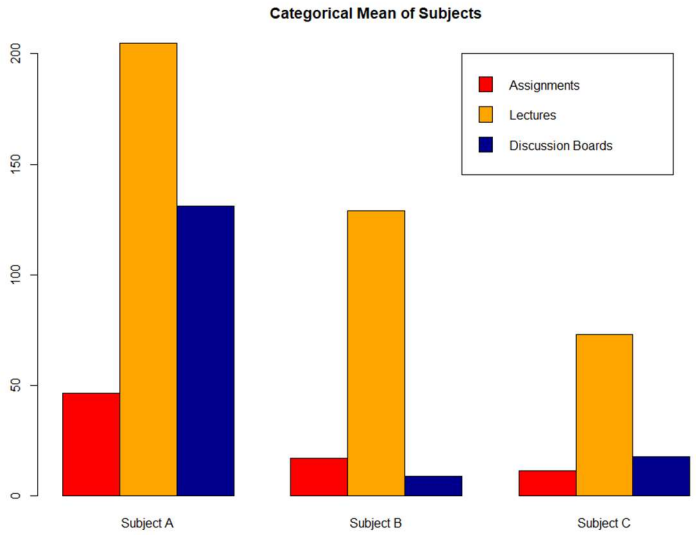


Figure 12 Average pageviews for 3 activities

The bar plot displays that Subject A has the highest views as compared to other 2 subjects which is probably influence by multiple reasons. Another observation is the relatively large difference between views of assignment and discussion topics of Subject A as compared to other subjects. Based on the plot, we can visualize the distribution of the variables. Overall, the distribution of lectures is obviously larger than assignments and discussion, and the distribution of discussion is mildly higher than the assignment. However, the result that we statistic from the data is a bit difference from the distribution of marking function. Here is the scoring standard given in Figure 13.

Criteria	Score	How to achieve the criteria (within week 0-3 of the semester) (Threshold Percentage: x = 100%)		
		Views x% of assessments	Views x% of Modules	Views x% of Discussion Forum posts and/or People Page
In addition to checking assessments and materials, examine how to connect with others	3	true	true	true
In addition to checking assessments, checks the materials (e.g., lectures and readings)	2	true	true	false
Checks only about major assessments and learning outcomes.	1	true	false	false
Insufficient evidence	0	false	false	false

Figure 13 Average pageviews for 3 activities

The marking function states that assignment take the significant influence for scoring. In other word, the necessary condition for scoring is to view the assessment. Furthermore, the distribution of discussion takes the least effect. Hence, there is a new research orientation, how to determine the distribution of each variable and what is scoring standard, which may be treat as the problem which we need to focus on the next semester.

4 Proposal for Next Semester

This section outlines the scope of the project, direction of execution and techniques or models used to implement the problem statement for the client. It also states plans for feedback integration and challenges that may arise during implementation.

4.1. Project Scope

- Analysis of the implementation and results of the current marking functions defined to measure self-regulated learning.
- Analysing the past and current student records to find correlations based on the marking functions provided.
- Extending the analysis through consideration of assignment grades to predict the scores of current students based on trends of the past students.
- For the initial phases, the data of students enrolled in a combination of subjects SWEN90009, SWEN90014, and SWEN90013 is to be considered to design a generalized model that can be applied to other subjects in future.

4.2. Semester 1 summary

After discussions with clients, we have shortlisted 3 subjects - SWEN90009, SWEN90014 and SWEN90013. We obtained the list of students enrolled in these subjects and identified the common students in all three subjects or combinations of two subjects. After analysing the marking functions provided by the client, we extracted the data required to calculate the scores using the Postman tool with the help of a token given by the client. This data was further analysed to find correlation between the variables of the items. Exploration of similar or related works under self-regulated learning was done to find models suitable for the obtained data.

4.3. Project Implementation

This part states the project implementation steps for next semester.

Phase 1

Phase 1 will involve comparative analysis of a single student's data across the 2 or 3 subjects they have enrolled in and finding trends in their behaviour based on the marking functions. It will also include analysing the abnormalities such as mismatch of the score generated by the marking function and score generated through self-assessment by the student. The score generated by marking functions will also be compared with the student's final academic grade of that subject to provide feedback to the marking functions. The linear regression model will be considered for fitting and predicting phase 1 data. The comparative analysis between actual scores and predicted score will act as feedback for the matrix as well as the model.

A challenge we might face while execution of this phase is that of underfitting the model to the data, due to the limited number of records.

Phase 2

Phase 2 will involve analysis of student data of all the students enrolled in a subject in a past semester. The past student data will be grouped into an ordered set of categories. The parameters considered, for past students will also be considered for current students for trend generation and pattern matching. This will allow us to generate suggestions for the students on their performance in case there is a need for improvement. Our predictions will be compared with their actual scores as feedback to our model. We plan to use the Hidden Markov model to recognize the behavioural patterns of students wrt LMS. The base for using this model has been provided in the related works section of the report.

A major challenge we might face is that of creating a bias with our suggestions to the students which might lead to deliberate behaviour to improve marking function-based scores.

Phase 2.5

Under this phase, we will consider the data of SWEN90009 for semester 1 of 2022. As this project has already been deployed, the students already have had the opportunity to score their work and have access to the scores generated by the computer as well. Both scores were generated using the marking functions. Using the analysis on same students in past semesters in phase 2, we aim to find if there has been any improvement in the behaviour of the students. This will help understand a bit more about the real impact of the customised dashboard in students' self-regulated learning behaviour.

Phase 3

This is a desirable outcome and is to be implemented only after completion of the previous phases.

In this phase we extract a list of keywords from the indicative content, ILOs and generic skills of the three subjects (using the handbooks) and match them against a dictionary of keywords extracted from job descriptions available on websites such as LinkedIn, Seek and Indeed to generate a general employability score of the subject. Further we could narrow it down to a single student taking an average score of all the subjects they have enrolled in to predict the employability score after taking combinations of certain subjects. This model can be extended to other subjects as well. This score will act as a reference to the university and will not be visible to the students.

Challenges:

- Constant update for the data on collected keywords from job sites.

Challenges in project:

The biggest challenge we expect to face in this project is the difference between the system generated scores predicting grades of the student and the actual grades of the student. In a scenario where a student just downloads the assignment document or the lectures and reads or watches them locally, the system will generate low scores, but the student may score good marks. Similar is the case where the student becomes aware of the criterion, and they modify their behaviour purposely to increase the scores. These occurrences cannot be accounted for but may majorly affect the outcomes of the project.

5. Timeline

5.1. Roles of group members

Roles	Vanshree Bapat	Wei Wang	Yixuan Liu	Simin Zuo	Jiayao Tong
Client communication and documentation	☑	☑			
Data collenction and preprocessing			☑	☑	☑
Data analysis	☑	☑			
Choose and build suitable model	☑	☑	☑	☑	☑
Deploy and validate model	☑	☑	☑	☑	☑
Literature review			☑	☑	☑
Team regulator		☑			

Figure 14 Tasks allocation for next semester

5.2. Project timeline:

Reviewing the tasks that we did in semester1, we have a primary understanding of this project after doing some preliminary analysis based on the data we extracted by the Postman. For the next semester, we dive deeper into the project.

For the first week, we will revisit the preliminary analysis done on the collected data and extract newly generated data till end of semester1 of 2022. In week 2 and 3, we will mainly focus on processing these data, transform them into usable format and start shaping the shortlisted models according to the methods mentioned in the related works. In week 4 and 5, we will analyse the chosen models and finalize the best fits. This phase (weeks 1-5) are the most important weeks of the project as by the end of it, the base structure for the project will be ready.

We will deploy and validate the models and consider the work of report in week 6 and week 7. And then, we will spend 2 weeks (Week 8 and 9) to communicate with client to demonstrate the findings but also improve our work based on the feedback given by the client. We have considered a buffer week i.e., week 10 which will allow us to catch up on any pending tasks. In the last two weeks, we will complete the report and presentations to finish the project. Figure 11 illustrates the plan for next semester.

Section		Duration
1	Section 1 - Data	2W
1.1	Revisit the preliminary analysis	
1.2	Extract newly generated data	
1.3	Data Preprocessing	
2	Section 2 - Model	5W
2.1	Start shaping the shortlisted models	
2.2	Analyse chosen models and finalize the best fits	
2.3	Deploy and validate the models	
3	Section 3 - Report	7W
3.1	Literature reivew and Introduction	
3.2	Methdology	
3.3	Result and Analysis	
3.4	Report conclusion and Preparation on presentation	
4	Section 4 - Feedback	3W
4.1	Demonstrate the finding to the client and request feedback	
4.2	Integrate client feedback into our work	
4.3	Buffer Week as an adjustment in case of delays or pending work from previous works.	

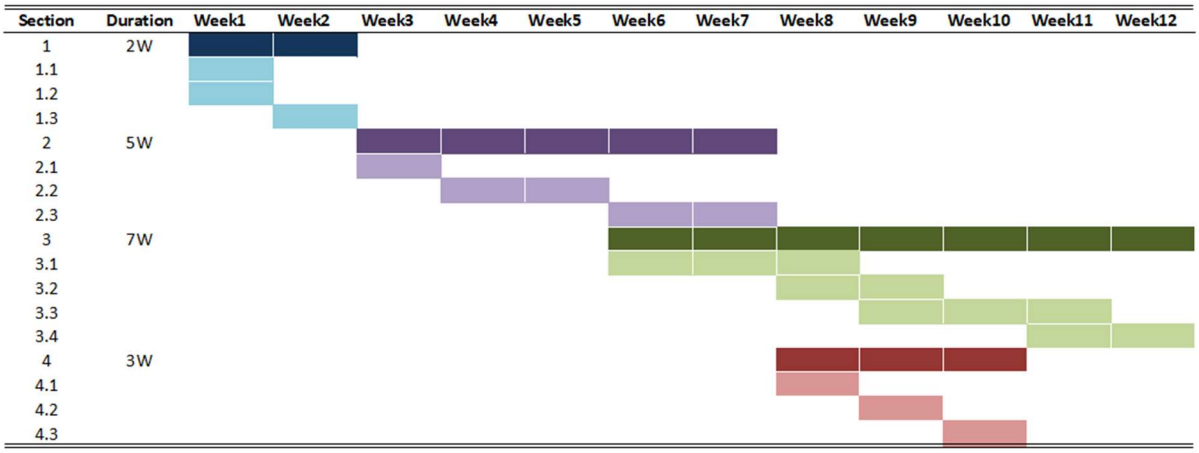


Figure 15 Timeline for next semester

6 Reference List

Abbasi, F. K., Ali, A., & Bibi, N. (2018). Analysis of skill gap for business graduates: managerial perspective from banking industry. *Education and Training*, *60*(4), 354–367. https://doi.org/10.1108/ET-08-2017-0120

Cosnefroy, L., Fenouillet, F., Mazé, C., & Bonnefoy, B. (2018). On the relationship between the forethought phase of self-regulated learning and self-regulation failure. *Educational Researc*, *28*(2), 329–348. https://discovery.ebsco.com/c/xppotz/viewer/pdf/naiof6vu6z

Fontaine, J. R. J. (2004). Exploratory Factor Analysis. *Encyclopedia of Social Measurement*, 803–813. https://doi.org/10.1016/B0-12-369398-5/00116-X

Hsu, A. J. C., Chen, M. Y.-C., & Shin, N.-F. (2021). From academic achievement to career development: does self-regulated learning matter? *International Journal for Educational and Vocational Guidance 2021*, 1–21. https://doi.org/10.1007/S10775-021-09486-Z

Khampirat, B. (2021). Relationships between ICT competencies related to work, self-esteem, and self-regulated learning with engineering competencies. *PLoS ONE*, *16*(12 December). https://doi.org/10.1371/journal.pone.0260659

Kitsantas, A., & Zimmerman, B. J. (2008). Self-regulation of motoric learning: A strategic cycle view. *Http://Dx.Doi.Org/10.1080/10413209808406390*, *10*(2), 220–239. https://doi.org/10.1080/10413209808406390

Klein, R. B. (2016). Principles and Practice of Structural Equation Modeling, Fourth Edition. In *The Guilford Press* (Vol. 8, Issue 5). The Guilford Press.

Lim, C. L., Jalil, H. A., Marof, A. M., & Saad, W. Z. (2020). Peer learning, self-regulated learning and academic achievement in blended learning courses: A structural equation modelling approach. *International Journal of Emerging Technologies in Learning*, *15*(3), 110–125. https://doi.org/10.3991/IJET.V15I03.12031

Lobos, K., Sáez-Delgado, F., Bruna, D., Cobo-Rendon, R., Díaz-Mujica, A., Hartley, K., & Andujar, A. (2021). *Design, Validity and Effect of an Intra-Curricular Program for Facilitating Self-Regulation of Learning Competences in University Students with the Support of the 4Planning App*. 11. https://doi.org/10.3390/educsci11080449

Maldonado-Mahauad, J., Perez-Sanagustin, M., & Beyle, C. (2020). A Questionnaire for Measuring Self-Regulated Learning in Massive Open Online Courses. *Proceedings - 2020 46th Latin American Computing Conference, CLEI 2020*, 400–409. https://doi.org/10.1109/CLEI52000.2020.00053

Mih, C., & Mih, V. (2011). Conceptual Maps as Mediators of Self-Regulated Learning. *Procedia - Social and Behavioral Sciences*, *29*, 390–395. https://doi.org/10.1016/J.SBSPRO.2011.11.254

Panadero, E., & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman‟s cyclical modelof self-regulatedlearning. *Anales de Psicologia*, *30*(2), 450–462. https://doi.org/10.6018/analesps.30.2.167221

Poon, J. (2014). Do real estate courses sufficiently develop graduates' employability skills? Perspectives from multiple stakeholders. *Education + Training*, *56*(6), 562–581. https://doi.org/10.1108/ET-06-2013-0074

PUUSTINEN, M., & PULKKINEN, L. (2001). Models of Self-regulated Learning: a review. *Scandinavian Journal of Educational Research*, *45*(3), 267–286. https://doi.org/10.1080/00313830120074206

Qiao, L., Zhao, W., & Xu, X. (2021). Mining and Analysis of Self-regulated Learning Process Model: Based on Hidden Markov Model. *Proceedings - 2021 10th International Conference of Educational Innovation through Technology, EITT 2021*, 276–281. https://doi.org/10.1109/EITT53287.2021.00061

Sarkar, M., Overton, T., Thompson, C. D., & Rayner, G. (2020). Academics' perspectives of the teaching and development of generic employability skills in science curricula. *Higher Education Research & Development*, *39*(2), 346–361. https://doi.org/10.1080/07294360.2019.1664998

Susaki, Y. (2021). Self-regulated learning and motor skills: effects of a physical education intervention program on Japanese college students. *Journal of Physical Education and Sport*, *21*(6), 3593–3598. https://doi.org/10.7752/JPES.2021.06485

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49. https://doi.org/10.1016/J.SYSTEM.2019.01.004