# Project Definition

## Intro

People who work within education understand the stress of teaching, they’re expected to work a full-time job and then also work within their own time providing feedback and marking homework. 17% of teachers spend over 11 hours a week marking homework [14]. As a result, there have been many approaches to reducing the workload: student self-marking, online homework websites [13] and plagiarism detection services (Turnitin).

Computer Science is still a new area in Education compared to old subjects like Maths and Science. Some places are still yet or only just adopting Computer Science now due to the UK governments financial boost [15]. Due to its infancy, there isn’t a perfect way to improve the way we mark coding homework. As a marker, you can’t always tell by reading code that its correct and that it will work, you must compile it or run tests on it to see if it works. Even this approach doesn’t always work, code that works on one PC may not work on another due to several variables, such as missing environment variables or different versions of a compiler. This makes marking more difficult compared to other subjects.

There have been attempts to fix this, there are sites like ‘Codewars’ that allow you to create unit tests for a coding challenge and then write code against that runs against those tests[16], but these require the marker to take the homework from one platform (that’s hosting the homework) to another (that tests or marks the homework). A good system will have all these requirements on one platform, which would remove the manual step in-between. Sites like ‘Codewars’ only allow autonomous marking through unit tests, which means a student is either right or wrong. An option to manually mark code will allow the professor to award points for things such as good coding practice.

Plagiarism has also been a big problem of Computer Science and checking the validity of coding assignments, there are services such as Turnitin, which search the web for similarities between two pieces of work. Problems arise when using this service for coding projects. Certain words and phrases are used to perform certain tasks and so as a result many pieces of code on the web are very similar. This can lead to work being marked as plagiarised when in fact, it’s just a common way of doing something. It’s possible to spot originality within code, this is because every programmer, develops their own coding styles. The clues to look out for are indentation, variable choice and commenting style, but these are hard for software to detect, but can be spotted when marked manually by a Professor.

This is an exciting project to me as it’s something that hasn’t been done before. Professors at Swansea University have been looking for a tool such as this and so making it a reality would be useful for the professors. This project gives me the opportunity to improve my skills and learn something new as I have no previous experience of working on a full stack web development project. These reasons and the fact that I get to make something that might be used after I’m gone is why I’m so heavily motivated for this project.

My goal is to create a web application that is a coursework management system like blackboard but also allows markers to review and mark coursework in the same web application by providing the same features web applications such as ‘Codewars’ provide [16]. The aim for this web application is to reduce the time it takes to mark coding course by reducing the number of steps involved and reducing the complexity of testing the code. Below are multiple features my web application will include that will achieve this aim:

* A professor will be able to set coursework and mark it within the same application.
* The web application will compile submitted coursework and run any unit tests the professor provides.
* The professor can also view the submitted coursework through an online editor that includes text highlighting and line indention.
* Each submission will get checked for similarities against other submissions within that coursework and notify the professor if similarities is above a certain threshold.

I believe that with good planning these aims are archivable. I have confidence in my abilities to complete a project of this size due to my experience of working in the software engineering industry and will be able to transfer the knowledge I have gained into my project. A large amount of the skills I have gained as a front-end web developer will also be transferable as this project requires a front-end too. Later, in this document I will go through my plan and provide evidence into achieving these aims.

Mark Scheme help

* The motivation and context are evidenced through reference to initiatives or reports from government, industry, public bodies or academic sources.
* There are clear, ambitious and detailed aims for the project. There is evidence they can be carried through to completion. Aims are imaginative and creative. These are conveyed precisely and effectively to the reader.
* Extensive independent research and effort using primary sources
* Mature and thorough evaluation of previous work including a critical comparison and reflection on the knowledge and evidence of integration into the project.

# Understanding the Topic

## Background Research

This section contains information about my project and the research into the structure, design and implementation of my project. As well as including relevant information about the range of different services and tools used within the project.

### Laravel

Laravel is a php web application framework, this is the main framework I will be using to build my project [6]. Laravel is an open source and contains many libraries, such as this one: [Laravel-Permission](https://github.com/spatie/laravel-permission), that can allow me to focus on implementing features like a complete permissions system [7].

During my research I have found plenty of free courses and tutorials [8] to teach me the basics of Laravel. I have learnt about useful features such as Migrations, which will help me develop and maintain a database. Authentication is another feature that will allow me to prevent unauthorised users from viewing certain webpages and manipulating the database [9].

Laravel requires other services in order to run correctly. A server is needed and so is a database framework. To solve this issue, I have decided to use Laravel’s homestead [10]. Homestead is a pre-packaged virtual box, which includes a web server called ‘Nginx’, ‘MySQL’ and more. This means the server used to run my project will be on a virtual machine. This means none of the requirements of the server will be part of my development machine. This requires me to have Virtual-Box and Vagrant installed on my development machine in order to run homestead.

### Use of Third-Party Applications

#### Java SE Runtime Environment

I plan to integrate the Java SE Runtime Environment 8 (JRE 8) into my project. This is a Java Virtual Machine (JVM) that will be obtained from Oracle and installed on the host machine. The JRE will be primarily used as a form of automatic marking of submissions. When a coursework deadline is reached, The JRE will be executed on the host machine and all the submissions will be compiled and if any Junit tests were found, they will also be run. This means when the Professor proceeds to mark the submission, they will know if the submission was able to compile and pass any provided tests. This is a form of automatic marking and can be used solely to mark submissions if set by the coursework creator.

The complexity of this is medium. The installation of the JRE on the host machine is simple. The path to the JRE will be added to the hosts external variables so can be accessed from the command line. The results of the compilation and tests as well as the output string from the compiler will be stored in a json string and will belong to a ‘submission’ on the database. This means the results from the JRE can be collected at any time and displayed on a webpage.

#### Moss

Moss is a third-party software tool used to detect plagiarism between files or programmes. This is different to services such as ‘Turnitin’ as it runs offline and only checks for plagiarism in files provided as input. Moss was created to be used in the classroom to check code was not copied in class. I plan to use Moss to check for similarities in different submissions within a single coursework. Its use will be to notify users marking the submission of similarities, not that submissions were copied or plagiarised, this will require manual investigation by the user reviewing the submission [2].

The complexity of this is medium. As Moss is just an application, it will be installed on the host machine. The path to the moss executable will be added to the hosts external variables so it can be run from the command line. Once the deadline of a piece of coursework has been reached, the Moss application will be executed and will check all submissions for similarity and store the result within a JSON string and will belong to a ‘submission’ on the database. This means results can be displayed on a webpage and viewed by an authorised user. The time it will take to run through all the submissions with Moss in big O notation will be: O(n2), as It will have to compare itself to every submission within the coursework.

#### Problems with invoking applications on the host machine

Running an application on a host machine will take a certain amount of time to complete. During this time more of the hosts machine’s CPU will be occupied by other tasks. This can lead to other server-side tasks taking longer to run while these applications are running.

If multiple coursework’s have the same deadline, this leads to a large server side slow down as multiple applications will be running on the server at the same time. This is because the applications are executed when the coursework deadline is reached. A solution to this can be creating a queue. So, the server-side applications only work on one coursework at a time even if they have the same deadline. This will prevent overloading the CPU on the host machine.

### Database Design

The database design is a key part of a web application. The front end’s primary use is to allow a user to view and edit the database, so getting the database correct first is very important as it will lead to less rework of code later. I will be using Artisan, the main tool for Laravel to create my database. Artisan contains commands that allow me to generate or delete my database and populate it with test data. This will be very useful when wanting to test by web application. It will also save me time and reduce several mistakes as I won’t be doing it manually and won’t be writing any SQL.

Converting these database drawings below to a database is simple in Laravel. Each table is a model class and contains the schema for a table. Laravel will create these tables using a command called migration, this is a source control approach to databases, which allows me to add schema to a database later or undo changes. I can then populate the tables I have migrated with Seeders and Factories. Factories is a way to create large amounts of fake data to go into a table, this fake data is then given to the Seeder which stores it in the database. This means if I was working as part of a team, I wouldn’t need to keep a database file in source control or transfer it to other developers as they can create their own data for the database with these Laravel features.

A screenshot of a cell phone

Description automatically generated

**Users**: A user is generic. A student and a professor will both be users. It’s the role within a module that makes has that provides them different actions compared to other users. A user can only have one role per module. A user can be assigned to many modules. A user can own many submissions.

**Modules**: A module can belong to many users. A module can have many coursework. A user with the correct global permissions will be able to create a module.

**Courseworks**: A coursework can belong to only one Module. A coursework can have many submissions.

**Submissions**: A submission can belong to only one Coursework. A submission is owned by a User. A submission will contain a JSON string. This string is in the JSON format, which means when its read, it can be converted to a JSON object and the information it holds can be read. Information contained inside the JSON string will include, compilation results, unit test results, in-line comments, and Moss comparisons.

**Roles**: A role contains a name and an ID. The name is to allow people to understand what the role is and does. An example of a role is student or Administrator.

**Permissions**: A permission allows the user to perform a certain action, such as edit coursework or create a module.

Below are the tables that will help complete the database relationships.

A screenshot of a social media post

Description automatically generated

Database Diagrams made with [dbdiagram.io](https://dbdiagram.io)

**Users\_Modules**: This table contains a list of all the users and the modules they are signed up to. This includes users with different types of permissions.

**Roles\_Permissions**: This table contains the relationship of roles and permissions. A user can’t just be given a permission. The user must be given a role that contains that permission. A role can have many permissions.

**Users\_Roles\_Module**: This table contains a list of all users with roles. Users are assigned roles within a module. A user who is a marker in one module may not have those permissions in another module.

**User\_Roles\_Global**: Users with roles within this table have permissions which can be used globally and not just for a specific module.

### Front End Development

#### Cascading Style Sheets

When it comes to styling the html for my webpages, I will use bootstrap as a framework. All the custom style sheets I write will complement this style. Bootstrap is an open source framework that contains great front-end component libraries such as navigation bars and containers. Using this framework will reduce the development time of the front end of my project and will reduce the size of my custom style sheets to improve readability. [4]

#### JavaScript

I will be using a JavaScript library to help speed up the development time of my front end. I will be using jQuery. This is a feature rich library which allows easy manipulation of html elements and contains useful JavaScript functions. jQuery is also open source and supports all modern browsers. [5]

### Testing

I will be using PHPUnit to write unit tests for my project. PHPUnit is perfect as it supports Laravel and allows me to test both my code (Unit Testing) and my models, controllers etc (Feature Testing) [12]. Laravel also offers commands to quickly make PHPUnit templates for you and it also includes a unit test file that offers Laravel help functions and enables the Laravel framework. This means that I can use these functions and reduce the amount of code I write as I don’t have to write some functions myself.

These helper functions can be used to test more complex things within the web application, these types of tests are called feature testing. For example, certain webpages can only be viewed if your logged in or contain certain permissions. The helper functions can be used to help recreate the exact same scenario for testing.

## Related Work

### Turnitin

Turnitin is the leading internet plagiarism detection service. Swansea University use this service to collect coursework submissions. Turnitin is integrated into blackboard and is primarily used for text matching and checking plagiarism within assignments that are submitted by students. Tutors are also able to use Turnitin to mark and provide feedback on student work [3].

Turnitin’s support for coding coursework is limited. It does not include line indention or syntax highlighting. Computer Scientists marking submissions usually download the files and view them in an IDE which includes these missing features **(PROVE THIS)**. My project relates to Turnitin as it’s an alternative way to specifically submit coding coursework and includes features such as syntax highlighting and line indention, so downloading files is not necessary.

The implementation of a similar feature to Turnitin, to search the web to check for plagiarism of submitted coursework, is out of the scope for my project. This feature is too large and beyond my knowledge and ability in implementation. As a compromise, I plan to implement a smaller feature that will check for familiarities between submissions for that specific coursework. I will achieve this by using third party software called ‘Moss’ [2]. This software checks for similarities between files.

### Blackboard

Blackboard is a web-based coursework management system. Swansea University uses this web application to manage coursework and modules. This is like my project as it’s also a coursework management system. The project will contain its own eco system of users, all with different level of permissions. Some users will be able to create Modules and coursework’s within them, whereas some users will be students to those coursework’s.

### Bitbucket

Bitbucket is like GitHub, it’s a web-based version control repository hosting service (I’m using this service to control my project). Bitbucket contains similar features that I want to implement within my project. When making a pull request on bitbucket, it gives other people the ability to review the code before it gets merged to a remote branch. Bitbucket allows users to highlight certain characters on a line and write comments of feedback for that line. This makes It very clear for people to understand the line of code the comment is referring to; I plan to implement a similar feature for my Project. When users receive feedback when using my project, they will be able to see which line of code the comment is referring to.

Mark Scheme help

* Demonstration of a sound body of understanding across the topic, and effective judgement in the selection of appropriate techniques.
* In depth understanding of technical background and evidence of ability to apply it in a constructive fashion.
* Critical evaluation of approaches and the current limits of knowledge are appreciated.
* Substantial progress has been made for this stage of the project.
* Original thinking has been demonstrated.
* Demonstration of conceptual understanding at the forefront of the discipline.

# Project Management

## Life Cycle

The model I have chosen to use for my project is the Iterative model. This model applies the waterfall model to multiple chunks of the project. This breaks up the project into something more manageable as it allows me to individually plan and implement each chunk separately.

As I have chosen this model, it means that if during the implementation of the project I discover an important software or user requirement that I missed, I can just begin documenting a new section and begin work on it. If I was using the waterfall model, I would not be able to do this as it wasn’t in the original software requirement documentation.

If I am unable to reach my deadline, I will still have some implemented and fully tested chunks, whereas if I used the full waterfall model, I may only have half implemented the project and not completed any testing.

I decided to choose this model over the V-shape model and the waterfall model as these are too strict when it comes to its different phases. I need the flexibility to keep moving through the phases of implementation and requirements as I am learning the Laravel Framework as I go. As I work on the project, I will learn more about the framework and its possibilities which could lead to me making changes to the implementation or requirements later into development. The Iterative model gives me this flexibility as it takes me back to the requirements phase after each chunk.

Front end development follows the iterative model nicely. Each webpage is a chunk, so I would design the webpage and create the requirements for it first, then implement it and finally test it.

## Plan

Working on multiple modules as well as my project will require great time management and organisation skills in order to complete this project before the Easter break. I have come up with ways of reducing the chances of overloading myself with too much work.

I have created a Trello board for my project ([Trello Board](https://trello.com/b/smZIFIKb/checkmate)) [11]. This allows me to break down my project into lots of different tickets. Some of the tickets are marked with ‘MVP’. This means that it’s a task that needs to be completed to achieve a minimal viable product (A product that works and is useable with minimum features). I will complete all these marked tickets before I move on to completing other tickets. This means I won’t be overloading my project with lots of features and means if I do run out of time, I still have a useable product to show for my efforts.

The amount of time suggested to spend on my project was around 300 hours, this is equivalent to 2 months of full-time work. The project is due in May and so from the first teaching week to the deadline is 32 weeks. This is roughly 9 hours a week I should be spending on my project. I have created a Gantt chart to help display the different areas of my project that I will be focusing my time on throughout the year. I have specified some deadlines throughout the year, these are goals that I want to meet for my project.

A screen shot of a building

Description automatically generated

The blocks represent a week of work. I have broken the workload up into three groups: planning, implementation and presentation.

The month of October will primarily be planning. I will be working on setting up my database and creating tasks on my Trello board [11]. I will be constantly learning Laravel and researching what it can do throughout October and throughout the implementation phase. I plan to have created all the front-end mock-ups before the beginning of my implementation. This will be used as a guide to make sure I include all features, but also to take usability and mobile devices into consideration when creating my web pages.

The implementation blocks contain arrows. These arrows show that I will be going back and forth from implementation to testing as I work my way through different chunks of my project. This relates to the life cycle model as I will complete a chunk of implementation for a page and then move onto the testing for that chunk before moving onto the next implementation. The implementation has been broken down into two subgroups. The first group contains the work required to complete the minimum viable product (MVP). The second subgroup contains the advanced features (ADV). These are features I will implement once MVP tasks are complete.

24/11/2019: At this time, I have projected to have prepared myself for Gregynog. This means producing a presentation and preparing a demonstration of my project. I would have completed most of my minimum viable project implementation and so will have something to demonstrate live.

April 2020: I plan to have completed all the implementation of my project. I would have prepared for the project fair and created a poster to display at my stand. I will also have made other preparations such as setting up another complete demo of my project. The exact date of the project fair has not been announced yet, so currently I plan based on the fair taking place sometime in April.

May 2020: By this time, I will have reached the deadline for my dissertation. I will have started work on my dissertation roughly 90 days previously. This is so I can slowly work on it overtime and leave myself plenty of time to make changes and make any required reworks.

## Risk Assessment

Risks are an obstacle that can prevent the success of a project. I believe it is important to assess the possible risks within a project ahead of time so that there can be enough plans made to prevent these risks or reduce their severity.   
I have created a risk assessment table [1] which identifies possible risks when completing different tasks. These risks are identified with actions to reduce their severity or prevent them entirely.

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Type | Hazard | Mitigation |
| Time Management | Generic | This risk emerges when failing to assess how long tasks will take to complete. This can lead to failing to complete work on time. This risk also emerges when working too many long hours which can lead to burn out. | A solution to reduce the severity of this risk is to follow the time management graph I have created. Writing down an estimate time to complete tasks will help organise my time. |
| Dependency Software Updates | Specific | Software projects have their dependencies. Using libraries that fulfil certain functions saves us time and maintenance. However, if a dependency a project uses updates and changes its behaviour it could break the project that depends on it. This can lead to a time-consuming process of trying to debug where the problem has come from which is expensive. | A solution to this risk is instead of just depending on the latest version of a software dependency, I will depend on certain version. This way I guarantee that the behaviour of that dependency will always be the same. |
| Hard Drive Error | Generic | It is possible a hard drive can become corrupt during development. This can lead to a loss of work on that machine, which can be very time consuming to replace. | A solution to this problem is to have the code base and other resources backed up by source control on a separate computer. This is so work is easily recoverable. I will be using A bitbucket cloud server as my source control. |
| Change in Development PC | Specific | Moving to a new computer for development can be stressful. The project compiles on the previous PC but may fails to compile on the new PC. This is usually due to a new development environment. Time will have to be spent setting up the new PC to match the development environment of the previous machine. | I will create batch or python scripts to set up the development environment on my PC automatically, this will make moving to a new PC easier. |
| Internet Access Issues | Generic | Development requires access to the internet. This is mainly required when fetching new dependencies from online, browsing the web for solutions to issues or committing changes to remote branches. The loss of internet can slow down progress on a project massively, as the tasks mentioned above cannot be completed. | I will make sure I am close to multiple locations that include access to the internet. I will use my Laptop as a development PC and so can move to another location with internet access if it is lost. |
| Git Server Failure | Specific | This project will be using a Bitbucket Cloud server to store its source code. It is possible that Bitbucket may go down for maintenance and so will mean that a lack of access to the source-controlled code. It’s also possible that this, wherever it is in the world could be damaged or stolen, this will lead to a loss of source code under source control which will be very time consuming to recover. | This risk is extremely unlikely, however a way to solve this is to keep a up to date checkout of my master branch on one of my local PC’s. This means if their server goes down, I have an up to date version on my local computer. |
| Requirement Overload | Generic | When writing the software requirements for this project it can be easy to overlook how long the features will take to complete. It’s possible that there will be too many requirements created with not enough time to complete all the requirements. | To avoid requirement overload, I have created a set of minimum requirements (Minimal Viable Product). These are the requirements that I will complete first before starting any additional work. |
| Feature Overload | Generic | It’s an easy mistake to focus on creating features rather than fixing bugs and writing tests. This can lead to a feature overload as the product may be feature complete but can be very unreliable and under tested which could lead to lots of bugs and issues not being discovered until deployment. | Feature overload is solved by the same problem above. The use of a Minimum viable product will make sure only the minimum required features are added first. |
| Inadequate User Interface | Generic | A user interface that is not easy to use will discourage new users to learn how to use the product and can lead to users being unaware of certain features and getting lost within the user interface. | To solve this risk, I have created some wireframes which focus on making a smooth and easy to use user experience. When creating the UI for the web application, these wireframes will be copied into the application. |
| Developing Wrong Purpose Functionality | Generic | When developing features, it’s possible that the point and use of that feature may be misunderstood. This can lead to a creation of a feature that does not fulfil the user requirements. This leads to wasted time and additional code being added to the code base. | To prevent wrong purpose functionality, I have created a set of requirements that have to be bet, I will not implement a feature if it does not reach any of the requirements. |

I have identified these risks by using multiple different approaches:

* I have investigated problems that have occurred for other previous third year students and industrial software development projects.
* I have investigated all aspects of my own project, including generic risks and specific risks related to my project.
* I have also spoken to other students with experience in software engineering as they may be aware of risks that aren't obvious to me.

I will continue to manage the risk of my project by following up with another risk assessment which will be used to identify new risks and update the risk assessment table above.

Mark Scheme help

* Effective judgement in the selection of a software development model has been demonstrated.
* There is a clear and precise description of how the chosen SDLC will be applied.
* The work schedule identifies deliverables that are detailed and ambitious.
* The size of each deliverable is appropriate and demonstrates progression.
* Each deliverable is associated with a method to measure its success, and risks are identified with mitigation.
* Delivery is timetabled appropriately.
* The reader is convinced that effective project management will take place as an integrated and useful component of the project.
* All risks are identified with mitigation. These include generic and project specific risks.
* It is clear how risks will be identified during the project, and risk management is a core component of the project.

Bibliography

1. <https://www.ccohs.ca/oshanswers/hsprograms/risk_assessment.html>
2. <http://theory.stanford.edu/~aiken/moss/>
3. <http://rref.henley.ac.uk/reading-cqsd/using-technology/technologies-services/cqsd-turnitin.aspx>
4. <https://getbootstrap.com/>
5. <https://jquery.com/>
6. <https://laravel.com/>
7. <https://github.com/spatie/laravel-permission>
8. <https://laracasts.com/series/laravel-6-from-scratch>
9. <https://laravel.com/docs/5.6/authentication>
10. <https://laravel.com/docs/6.x/homestead>
11. <https://trello.com/b/smZIFIKb/checkmate>
12. <https://www.cloudways.com/blog/laravel-unit-testing/>
13. <https://www.mymaths.co.uk/>
14. <https://www.tes.com/news/workload-tens-thousands-teachers-spend-more-11-hours-marking-every-week>
15. <https://www.gov.uk/government/news/schools-minister-announces-boost-to-computer-science-teaching>
16. [https://www.codewars.com](https://www.codewars.com/)