Exploring the Usability of Terrain Generation Tools Through an Iterative Design Process

GDEV60001 GAMES DEVELOPMENT PROJECT

Cameron Dawes

SUPERVISOR: BENJAMIN WILLIAMS

SECOND SUPERVISOR: JAMES VICKERS

Contents

Abstract

Development tools frequently struggle with a functionality first ideology resulting in the use of tools feeling challenging. Past research indicates developers at times lack the resources to consider usability as a foremost priority. Terrain generation tools are one such example, as complexity is ever increasing due to the reliance on its capabilities and the abundant research in the field. This complexity is exacerbated by the lack of research that focuses on the usability of terrain generation tools and development tools as a whole. The aim of this study was to evaluate the benefits to user experience from consistent implementation of results gathered from usability testing and feedback throughout iterative development. A terrain generation tool was developed in order to gather usability data and feedback through the completion of tasks which are indicative of any usability shortcomings and frustrations users experienced. This was then utilised to improve the tool in the following iterations and tested again. Within each feedback phase the System Usability Scale questionnaire was used to evaluate the tools overall usability. Upon the completion of testing, the SUS scores for each week were compared, with an overall increase of usability from the first iteration. The score dropped within the first two iterations; conclusions can be drawn from the data that the causation was the introduction of tasks which prompted the participant to engage more deeply with the tool giving a higher accuracy of feedback, it can be inferred that the first iterations score may have been influenced if tasks were present. Further research could aid the development process as it could be narrowed to a more proven method instead of a combination of different methodologies. In addition, developers can infer from the research that usability findings can be directly integrated to improve user experience.

Introduction

This dissertation explores the usability of terrain generation tools through an iterative design process supported through consistent usability testing and feedback.

Development tools are a core aspect of development, allowing for streamlining of development pipelines. Terrain generation has a large body of research which is shown in the continuous advances within the field; and as such terrain generations tools for games and other media platforms have integrated more functionality and features into their tools which in turn has caused the complexity to increase. This increased complexity influences the tools usability and learnability, additionally research is inadequate in the usability and learnability of terrain generations tools and development tools in contrast to work done advancing terrain generation in general. Within the field of software and tools development, strides have been made in the advancement of usability although in large organisations and projects usability is often disregarded in preference to its functionality as found by Patel et al (2020). Furthermore, many development tools are targeted at experienced users with their vast functionality making them difficult for novice users to use or adopt.

Due to the lack of research in the usability of development tools as a whole and even more so in regard to terrain generation there is little data on balancing the tools complexity with its usability. Improving the terrain generation tools usability could result in a more efficient and usable experience for users along with making it accessible to a wider range of users. In addition, the focus on usability could make the work flow of the tool more efficient which would result in an increase in productivity.

This study will analyse existing research on usability and development tools. Along with identifying key changes throughout development through the use of usability testing and the SUS questionnaire resulting in suggestions for improved design for general and terrain generation tools.

Aims and Objectives

The aims of this study are to investigate the current state of literature regarding the usability of games development tools as well as the identification of the core aspects that affect the usability of a terrain generation tool. This study will concentrate on how usability focused design decisions impact the efficiency, learnability and usability in terrain generation tools and development tools.

In general, this dissertation will consider the effects of consistent implementation and consideration of user feedback from usability testing, evaluating the benefits to the user experience and thereby provide a recommendation to developers of how to consider implementation and use of usability testing.

Dissertation Structure

This dissertation is structured to explore the impact of usability-focused development in terrain generation tools. Firstly, a review of current literature on usability of software and development tools will be conducted to identify challenges, short comings and findings that can be used in the studies development process. Secondly, planning and development will start where surveys will be completed to gather user data predominantly focusing on usability and general improvements. Utilising the feedback, further development will be completed to improve the usability and functionality of the tool, this iteration will be tested and its usability measured using the SUS questionnaire alongside additional qualitative and quantitative feedback which used in further development. Finally, the results will be compared to determine whether the focus on usability resulted in a more efficient and usable tool.

Literature Review

This literature review aims to examine the existing research of usability in software, usability in tools, game development tools and the focus of usability in game tool development. Firstly, it evaluates the identification of a lack of usability in software development. Secondly, designing the usability of tools and software. Thirdly, the evaluation of usability in software. Finally, the evaluation of the closest research, consisting of two studies which focus on developing and evaluating usability in game development tools. Whilst a lot of research has been done on terrain generation and development tools, there is a lack of studies focusing on their usability during development and its evaluations.

Patel *et al.* (2020) present a concern regarding the lack of expertise and "capacity to address accessibility" in software. Consequently, a large quantity of software is published with little consideration or time given to software accessibility. This paper investigates the challenges technology professionals face in including accessibility and design in the development process. The findings from the study conclusively found that the majority of participants in the survey had not learnt anything substantial and a participant stated, "Like if I were to think about how I applied what I learned from… [my institution] about accessibility I don't personally feel I'd be confident in applying what I did learn in those courses like in the real world." Indicating that the already small size of people that were taught accessibility do not have the confidence to put it into practice. The assumption can then be made that most tools in the game development space suffer from the same lack of accessibility considerations and so action needs to be taken to consider them.

Putnam, *et al.* (2012) further explored the inclusion of accessibility considerations by UX/HCI professionals into their work. Results showed that 70% of the participants had some sort of consideration for accessibility whilst 19% of the participants said they should consider it and the remaining 11% had no consideration. This data seems rather inspiring but not every project, especially ones that are made by smaller teams have dedicated UX/HCI professionals and such the interface is left up to the developers. This may lead to improper implementation of accessibility features and the development of the tool lacking consideration for accessibility. When an attempt to implement these features are done, they are to a mostly finished project in contrast to the better practice of implementing and designing for accessibility alongside development. Putnam *et al.* (2012) concluded that "Considering accessibility and diverse users in HCI/UX professions is (1) a good decision on moral, financial and legal grounds and (2) well supported by organizations like the W3C." and that many existing software are not accessible. They finish with emphasising the importance of educational programs in IT that prepare students to consider and advocate for inclusive design.

Borg, *et al.* (2020) studied the participants of the Global Game Jam. A Game Jam is a challenge to develop a game based on a given genre/topic in a short time span. The survey investigated the number of years of experience in game development of the participants. Evaluating the results they identified that that the majority of participants had between 1 - 4 years of experience followed by the second largest group with <1 years of experience. In addition, a large portion of solo participants reported being software developers which proved to be true for teams as well. From this we can conclude that the ability and expertise of the developers were narrow, which leaves a shortage of ability in art, modelling and sound fields. These findings support the creation of simplified development tools focused on making the design processes more accessible. A simplified terrain generation tool aimed at being easy to learn and use would fit the narrower functionality needed.

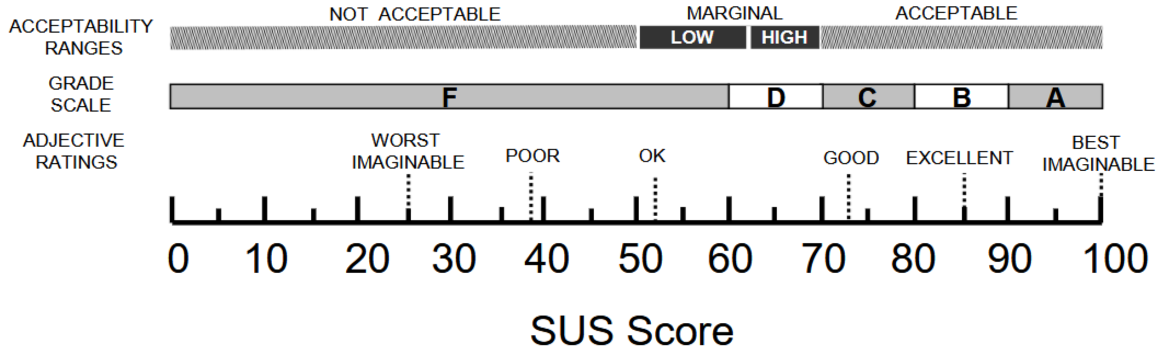
In the book 'Designing the User Experience of Game Development Tools' written by Lightbown (2015) highlights the methodologies for developing a tool's UX. Lightbown describes the user experience of a tool as a three-layer pyramid, bottom up it is: Useful, Usable, and Desirable. In addition, it is discussed that a large portion of tools meet only the bare minimum of the pyramid, being useful, which leads users to feel forced into using the tool even if it is challenging. Lightbown emphasises a focus on user centred design, this relies on the identification of a target group or the largest user body and focusing development efforts on those needs. A great way for developing usability is watching a user utilise the tool and observing their use of features, it can be observed that some tools require workarounds to function correctly. This indicates areas of improvement for the tool which could not otherwise be easily identified through qualitative surveys.

Avouris (2001) in his workshop on software useability delves into the core aspects of usability in software and explains standards that must be met, these consist of “Understandability, Learnability, Operability, Attractiveness and Compliance”. Following this is a discussion on measuring usability, a popular method is splitting a user’s feedback and performance into three categories, effectiveness, efficiency and satisfaction. Effectiveness looks into the amount of achieved goals, how many users successfully completed tasks they set out for and average accuracy of the completed tasks. Efficiency is measured by time to complete tasks and an average task completed per measurement. Finally, satisfaction is measured by a rating scale, frequency of discretion and frequency of complaints, these factors of usability can be seen within the standardized questionnaire PSSUQ an alternative to SUS though SUS still focuses on the same three areas: effectiveness efficiency and satisfaction.

Weber, Zoitl and Hußmann (2019) argue that the practice of model driven development which uses tools to “support and empower developers” are actively held back by the lack of usability as it acts as an obstacle for the adoption by novices and experts alike. Weber states a tool can only be successful “by being usable, supportive of their users’ goals, and by facilitating learning can they fulfil their purpose.” Subsequently, the issue was addressed within the usability research and open-source space where they fall short with regards to usability. Furthermore, this lack potentially leads to a reduction in the productivity and/or comprehension of the tool and in its worst case “results in users simply not using the tool, which not only means a lot of development effort gone to waste but also stops users from benefiting from the gains of MDE, even if they are aware of the theoretical benefits”. The study concluded the existence of “an abundance of usability issues” where they tackled them with the feedback from qualified test participants. It was mentioned that their efforts were met with appreciation and engagement in the development and improvement of the accessibility.

Usability testing is at the core of most related research, usability questionnaires are used in conjunction with testing to quantify the tools usability. The System Usability Scale is the most widely used standardised usability questionnaire developed by Brooke (1995). The SUS was developed to take quick measurements of how people perceived the usability of computer systems. Brooke states in his 2013 retrospective that the SUS proved to be an extremely simple and reliable tool for usability evaluations and was widely used as a standard even though it was never formally standardized. He then discussed how to get non bias responses changing the tone of the statements from negative to positive “Respondents read each statement and make an effort to think whether they agreed or disagreed with it.” A benefit of SUS compared to some other questionnaires for this researches purpose is that it has been utilised for more than 3 decades and as such has been studied and analysed on its performance and the evaluation of its data. Brooke discuses a paper analysing around 3500 SUS results performed by Bangor, Kortum and Miller (2009) which resulted in a decided score rating system which added adjective ratings to the SUS score.

Figure 1: A comparison of the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score



(Source: Journal of Usability Studies, Bangor, A., Kortum, A., Miller, J., 2009, p. 121).

In a comparative analysis between game development engines by Christopoulou and Xinogalos (2017) they identify the growing broadness of the population who use game development engines saying, "People involved in the design and development of serious games come from various fields. For example, pedagogists and domain experts with limited, if any, coding skills are involved in designing and developing serious games". This indicates a need for development software and tools that are accessible for individuals without former programming knowledge. This is discussed in their conclusion, where overall unity is a more accessible and beginner friendly software. Included is a requirement of C# experience for use compared to Unreal engine 4 which does not require any programming experience, it uses visual scripting but overall is a more complex tool and far more hardware demanding than Unity and other existing software. In their analysis of different game development engines accessibility, it was identified that Unity "Seems to be the most usable game engine, providing the most free tutorials, examples and assets, while its community is very large". Unity's lack of technical support however and that it is closed source compared to other mainstream game development engines such as Unreal Engine and Godot, make developing tools for the engine complex. It is ultimately concluded that there is no best engine for development as they both have pros and cons, which is the consensus of other research focusing on the comparative analysis between engines (Vohera, C. *et al*, 2021).

A paper by Toftedahl and Engstrom (2019) initially set out to evaluate existing localization features within game engines but identified the complexity and interconnection between game engines and tools, which made researching specific functions within the engine difficult. This resulted in a change for the direction of the paper to a "Taxonomy of game engines and the tools that drives game production." In this paper they identify that tools are a core part of the development pipeline; they supply a definition for "User facing tool" that are "designed to support human developers to create game content". Furthermore, they heavily emphasise the "strong focus on usability" which is required for tools in the game development space especially due to the wide use of development tools in every part of the pipeline.

In Kasurinen, Strandénand and Smolander (2013) study on "The developers' expectations from development and design tools", the aim was to understand the expectations attached to the tools used in game development practice. They investigated new startups and established studios. Research contributed to the understanding that game development tools are primarily used to test concepts and allow for prototyping. Additionally, the motivation for these tools' creation was related to the number of bugs in existing tools or alternatively an unintuitive user interface design. These tools are expected to be able to adapt to changes during the development process. Using these insights, the terrain generation tool targets inexperienced users rather than established studios, as studios are content with their existing more complex tools. By prioritizing accessibility and a usable interface, the tool better supports testing and creation for users who may struggle with traditional development software.

Volioti *et al.* (2024) introduce Codeless3D, a low-code tool designed to enable individuals without programming expertise to create 3D game environments. The motivation for development was the identification of the ever-diversifying teams at game development studios. The increasing specialisation of members causes a bottle neck as designers and non-code developers rely on programmers to test feature implementations. The main goal of the tool is to reduce design and development members reliance on programmers, enabling developers to test and iterate independently of other members of the team. To assess the success of the tool the SUS score was used alongside qualitative data to evaluate the usability of the tool and to gather developer's opinion on its effectiveness. It was concluded that the wider industry could benefit from such tool streamlining the design and development process whilst making it more accessible to non-programmers.

Blomqvist and Detterfelt (2020) discussed the challenges developers face with "Usability and quality of life" of the tools they use. The research discusses a lack of usability in tools which leads to frustration especially when used in "crunch periods". It was stated that this can be avoided by "creating effective, functional and user-friendly" development tools. In addition, the tool was developed using an iterative development methodology with user testing after each iteration to identify any usability issues. It was concluded that the use of consistent testing and user feedback resulted in a tool that can be "considered highly usable in terms of effectiveness and averagely usable in terms of learnability". This was determined using the SUS questionnaire in their final development cycle. Throughout development the tasks were utilised to prompt the user into fully experiencing the tool. Participants were evaluated by their progression through the tasks. Testing was conducted in person which allowed for insights to be gathered by the researcher in real time.

The reviewed literature highlights the large role of development tools in games and software development and the gap for attention given to its usability. This study aims to expand the understanding of the benefits from usability testing for terrain generation tool development beyond the current research. In addition, the findings could be used to further sequential studies which focus on the evaluation of the usability focused development process and the positive impact it has.

Development Methodologies

For the development of the terrain generation tool an iterative design process was followed as it enables the examination of meaningful insights with respect to a tools design and usability, this design process is commonplace in existing literature, (Dow *et al*., 2005; Følstad, 2017). Savage (1996) successfully used the iterative design process in their research identifying users heavily preferred one interface over the alternative. This section outlines how the development and research methodologies were integrated to accommodate the user needs and was refined based on the feedback received.

A range of development methodologies were considered. Despa (2014) undertook a comparative study of 21 software development methodologies. The five methodologies considered for this study were follows:

Prototyping
The process of developing a sample product that is then tested and given feedback on by the end-user, iterated and repeated until the end-users are happy, and the development of the actual software begins. In this process the participants would give feedback, it would be evaluated and used to further develop the prototype to meet the specification and then repeated. The dilemma faced is that a prototype is developed to meet the bare minimum criteria for the specification and not developed with the intention of being the final product until the user decides it is completed and then it enters the actual development phase and thus would not be compatible with the goal of releasing a usable tool.

Iterative and Incremental Methodologies

Unlike Prototyping, this approach builds upon previous iterations rather than discarding them, reducing wasted development. This methodology accommodates the limitations of a one-person team along with the time given between iterations to receive feedback. This was not suitable for the research due to the time constraints of development and accommodating a longer period for data collection.

Rapid Application Development

This development methodology splits iterations into modules. Feedback is given on separate modules and can be developed simultaneously; iterations of modules can start the second feedback is received on it. This approach is incompatible with the development timeline of this research and the ability to get user feedback as it is unsustainable to receive feedback on each individual module.

Scrum

A sprint-based methodology where each sprint results in a functional prototype. Development starts with a backlog of tasks; each sprint should last no longer than 4 weeks. This aligned well with the approach intended for this research where each version of the tool is rapidly released for testing in short intervals. The issue with this methodology is that the back log must be finalised prior to the start of the next sprint which results in a lot of misused time waiting on feedback.

Adaptive Software Development Methodology

This final methodology is like Scrum though it differs as it accepts feedback in all stages of development compared to getting feedback at the end of a sprint which is more constraining. Comparing this to the other methodologies it is clear that it offers the structure of set development whilst offering the flexibility of adding the feedback from the surveys throughout development.

The chosen development methodology was Adaptive Software Development due to the time span, quick iterations and constant development throughout user testing. Further benefits of iterative development include the ability to collect meaningful design suggestions from participants and apply them during development, compared to the alternative, non-iterative development which surveys participants a single time at the end of development, it is determined that "involving the users as testers improves the design cycle as well as the actual design" (Hellstén, 2019, p.36-37). Iterative design enables the incremental improvements of prototypes to achieve the design goals for the project (Sharp and Macklin, 2016, p124).

Before the research was conducted a minimum viable tool had to be made and the initial scope of the tool had to be determined. During development a weighted backlog system was used to keep track of the features and changes that had to be made within the current development cycle.

Research Methodologies

Understanding how users interact with development tools is at the core of improving its usability. In order to achieve this a research method was required to gather this data. A mixed methods approach aligns with the objectives of this dissertation in exploring the usability of terrain generation tools. Qualitative data enables deeper evaluation of user preference and allows for the extraction of greater insight from participant feedback. On the other hand, quantitative data offers more concrete metrics on the tool's usability, for example through the SUS. The combination of both offers a diverse approach to evaluating the usability of designed tools (Venkatesh, V., Brown, S.A. and Bala, H., 2013; Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A., 2007).

A core method for quantifying the tools usability had to be selected, there are two major classifications of usability questionnaires: post-study(post-test) and post-task. The differences between these classifications are that post-study questionnaires are to be filled out after all tasks have been completed whilst a post-task questionnaire is to be filled out after each specific task. After investigation, post-study questionnaires were chosen due to the less intensive nature of the questionnaire, along with the goal to evaluate the tool instead of targeting specific aspects. A selection of four questionnaires were considered: the System Usability Scale (SUS), the Post-Study System Usability Questionnaire (PSSUQ), the Computer System Usability Questionnaire (CSUQ) and the Software Usability Measurement Inventory (SUMI) (Brooke, 1995; Kirakowski and Corbett 1993; Lewis, 1992). Each of the questionnaires have their strengths and is designed for slightly different usability related analysis.

Three different scales were evaluated for this research project. Firstly PSSUQ, developed at IBM, is a 16-item questionnaire with 7-point Likert scale along with an N.A. choice, often used as an alternative to the SUS, this survey gives more targeted scores (Lewis, 1992). The score is calculated by averaging the values between 1 (strongly agree) and 7 (strongly disagree) of the 16 questions to represent the overall usability of the software. It has three subcategories, System Usefulness, Information Quality and Interface Quality, these three subcategories are averages of specific numbered questions. According to Lewis and Sauro (2009) the PSSUQ is as successful for measuring usability compared to the SUS although its usage is far less at 15% when compared to the SUS's 43%.

Secondly, SUMI is a 50-item questionnaire, with three choices, Agree, Undecided and Disagree. Along with this, the SUMI questionnaire requires the use of a license making it unsuitable for this research.

The SUS questionnaire was the chosen usability survey, it's a 10-item questionnaire with 5-point Likert scales. It was chosen due to its ease of integration, use and analysis of the results as they can be combined into a single SUS score. Additionally, the SUS questionnaire offers more comparative data due to it being more frequently used which can strengthen the validity of the score achieved (Lewis and Sauro, 2009).

An additional reason for the choice of the SUS questionnaire is the large amount of data backing the validity of the SUS and the plethora of studies validating the accuracy of the questionnaire. Bangor, Kortum, and Miller (2008) conducted a study on the validity of SUS results which shows that the correlation between the results of an SUS study and a 7-point scale also used to measure usability had a "strong correlation ($r_s = 0.525$, $p<0.01$)". Bangor, Kortum, and Miller (2009) conducted further studies on the SUS comparing results between gender, age, technical ability, along with adjectives by changing the 5-point scale on the SUS to a 7 point scale using Worst Imaginable, Awful, Poor, OK, Good, Excellent and Best Imaginable shown in Figure 1. When the survey with changed adjectives was compared to the original survey, results were strongly correlated. In cases where dependent variables were measured such as a user's perceived technical ability their SUS score had a strong correlation. In addition, the SUS questionnaire is easy and quick for the user to fill out while maintaining high accuracy results. Furthermore, research shows that the SUS score doesn't fall short by recurring participants as much as other tools might and does not have a bias in wording like some surveys Bangor, Kortum and Millter (2008). Finally, the SUS Score was used in the two adjacent studies creating and evaluating developments tools, Codeless3D by Volioti *et al* (2024), and the Real Time Integrated Tools by Blomqvist, Detterfelt (2020). Sharing the SUS score with these two closely related studies enables the ability to comparatively evaluate the results of the study and come to a conclusion not limited exclusively to this research.

These two studies differed in their use of the SUS questionnaire, Violioti *et al* (2024) used a single survey at the end of the study on the finished prototype while Blomqvist and Detterfelt (2020) use it throughout development as they iterated the tool. The project by Violioti *et al* had limitations in some areas as they were unable to gather suggestions during testing, leading to a lack of certain features that could have benefited the final product. My methodology closely mirrors that of Blomqvist and Detterfelt, after evaluating the separate data sets for both studies it is evident there is more to gain from consistent testing, when compared to a single test. In addition, consistent testing allows for greater amounts of data which can be used to further support the validity of user testing during development.

In addition to the SUS questionnaire, Tasks were used to quantify how easy the tool is to use proficiently, furthermore, its application guides the user to utilise all aspects of the tool which improves the breadth of feature testing and bug finding. These tasks followed the guide Moran (2018) has for writing tasks for usability studies, ensuring clear, unbiased instructions and a single method of completion. Part of this guide emphasises providing as much detail as needed in order to keep the task narrow and focused, each task should stand alone. Data was gathered through two questions, three for the final survey focussed on the confidence of the participants prior to completion and the task duration. Three of the four surveys included tasks, as the first tool released lacked the complexity needed to utilise tasks to guide the user to use the tool fully.

Along with the SUS questionnaire, which is the core of the survey, participants reported self-metrics through including the participants perceived computing power in a Likert scales from very low to very high. This data was used alongside later questions which asked for the tool's performance, causes of the performance and comparison to previous iterations to allow for identification of areas needing improvements.

During development a tutorial was created to be supplied alongside the tool, due to this a question was added in Round 3, asking the participants how effective they found the alongside taking feedback on how the tutorial could possibly be improved upon.

Following the SUS questionnaire, a question for recurring participants asks in comparison to the previous tool how the usability and functionality has changed from much worse to much better. Along with that, a qualitative question was put for participants to suggest changes to the UI layout which they would like to see in the next iteration.

The survey asks the participant how long it took for them to feel as if they understood all the functions of the tool, Immediately, after a couple minutes, after an extended period of time and You didn't understand the tool, using this along with the task data and whether people are recurring participants we can evaluate the first impressions of the tool.

The final three questions focus on the tool's features, "were there frustrations in the tool and how would the participant want them improved", "where their bugs using the tool and what were they", and "are there features they would like to see added to the tool that does not already exist".

A total of four design sessions were run throughout the study. Each of these were iterative, following on from feedback from the previous session. The feedback given by participants in one session was utilised to adapt the tool for the next. New and recurring participants experienced each of the four sessions. Recurring participants enable a consistent thread of feedback and yield cross-exposure insights whilst new participants offered unique perspectives and control for bias regarding previous exposures. After each session, the post-study questionnaire was filled out by participants. The study was conducted online, and participants were given access to the information form which explained the study's purpose, the website provided instructions on the use of the tool and the tasks along with providing links to all necessary web pages.

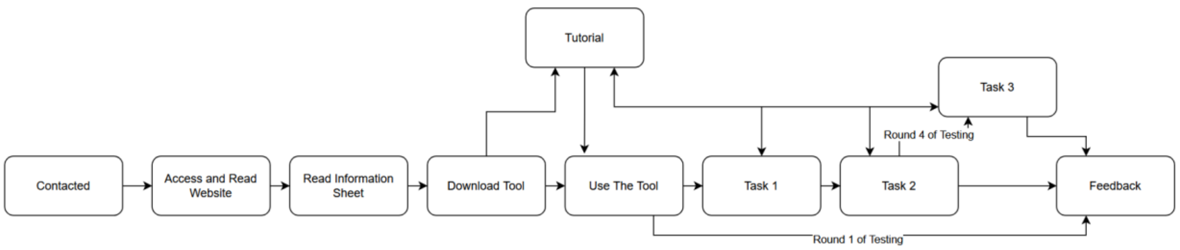
Participants

The goal of the study is to focus on users unfamiliar with contemporary terrain generation tools. Consequently, the recruitment of participants was indiscriminate and 600+ possible participants were contacted through an online campaign advertising the study. For the final survey a lack of participants were identified so to recruit more participants demonstrations were given in lectures within the University using the tool accompanied by a request for people to participate. The study was granted ethical approval subject to the ethics board of the institution and no reimbursement or incentives were used to recruit participants.

Participants went through below path of participation:

- They were contacted about the study,
- If they decided to participate, they proceeded to the website
- They are supplied the information sheet
- They download the tool and read the instructions
- From week three and beyond they can look at the tutorial
- From week two and beyond they will do the tasks, in week four there was an additional task
- After completion they fill in the survey

Figure 2: Path taken by participants



Materials

Throughout the study, a multitude of third-party tools and libraries were used.

- **Github -** a cloud code/file hosting tool, I utilised this to both save iterations of the tool throughout development along with hosting the executable files for the tool so that participants could download it from there if it was preferred. It also acts as a website host which is where the information for the study was held and distributed to participants (Cameron-cmd.github.io).
- **Microsoft Forms -** the chosen method for collecting data from the participants as it was recommended by the university.
- **OneDrive -** a cloud-based file storage provider, this was used to host the alternative file location for the tool giving the users the ability to choose by preference. OneDrive allows you to make the files accessible by a shareable link.
- **Visual Studio 2022 -** the chosen programming software, all programming was done within it.
- **RenderDoc** - a debugging tool for graphics based programs and allows the user to see what is happening on the graphics card during run time.

Within the tool seven libraries were used and one project was referenced:

- **Tinygltf -** a small library for creating and reading GLB and GLTF files.

- **ImGui -** a public GitHub library was chosen to be used for the tools UI as it is widely used, documented and tested along with the ease of implementation and alteration of created elements.
- **FastNoiseLite -** a public GitHub repository which creates a multitude of noise efficiently making it suit the project well.
- **DirectXTex** - a low code texture processing library for reading images into DirectX11 developed by Microsoft
- **DirectX11 -** created by Microsoft is a rendering and physics pipeline for computer rendering.
- **NFD (Native File Dialog) -**allows users to directly interact with windows built in file explorer.
- **Hydraulic erosions algorithm - a**an efficient simulation of hydraulic erosion developed by Nick McDonald (2020) which was adapted for the project.

Procedure

Starting the project, the goal was to have development span roughly 5 weeks, sprints were to take no longer than a week and half. The first development cycle required a minimum viable tool that had to be created, this tool needed the ability to generate terrain and export it. After the first version had been completed the survey had to be made and sent out. This was created including the SUS questionnaire as it was a core part of the research along with additional qualitative and quantitative questions to assist with the development of the tool's usability. A website was subsequently developed using basic HTML, this site was used as the hub for all the information relating to the study, this was done to simplify the process of onboarding participants by requiring only a link to be provided which would then give them access to everything they would need, the information sheet, my contact information, the two download locations, along with some guidance on the tool and the link to the form.

After everything was prepared, the link was sent out in both public forums and directly to students and lecturers via teams accompanied by an explanation of the study and request to participate. Development started whilst the survey was still being completed by participants to optimise the development cycle, once feedback had been gathered, it was analysed and used to further develop the tool.

After the second round of development, adjustments were made to the website and survey to accommodate the integration of task into the testing, this was to guide participants to use the full functionality of the tool, the tasks were listed on the website and four questions were added to the survey to record the users experience with the task. Additionally, there were questions added to allow recurring participants to give feedback in comparison to the previous iteration, this allows analysis on the perceived improvements by users.

In the third round of testing a tutorial was made to accompany the tool, this tutorial was linked on the website along with questions on the survey on the effectiveness of the tutorial.

In the final week, a third task was added as the two tasks were not enough to include all the functionality of the tool. Which resulted in two additional questions in the survey.

Implementation of Material

Throughout development, debugging was a core part of getting features to work, one limitation with the graphic based system was that debugging the program was difficult due to visual studio not natively supporting graphics debugging, instead RenderDoc was used as a graphics debugger tool which lets you capture a frame and walk through the process step by step.

A large part of the tool was the UI and a simple straightforward to use UI library called ImGui was implemented. In the first version of the tool noise generation was implemented, this is a complex algorithm and so a FastNoiseLite, a library, was used for its generation. Along with this a hydraulic erosion algorithm was adapted from Nick Mcdonalds.

In sprint three due to the request of users, integration with windows file explorer was implemented using Native File Dialog, which is a file access system which supports Windows, Mac and Linux.

In the final sprint, the file type GLB was determined to have the widest adoption and application and was used, Tinygltf is a library that focuses on streamlining the process of reading and writing GLB and GLTF files.

Results and Findings

Sprint 1

After identifying the minimum requirements for the tool to be usable Sprint 1 commenced. The minimum requirements for the tool for the first round were determined to be the functionality needed for it to be a usable tool which meant it had to be able to generate terrain along with exporting the terrain into a usable format which for the first sprint was a .obj file. As shown in Figure 3, the sprint focused on the implementation of a viewable terrain within the window alongside basic UI elements, terrain generation features, Camera movement and the ability to export the model. These features were recorded in the backlog for sprint 1 which can be seen in table 1.

In this first version of the tool there are three main UI elements: the top left element which incorporated the diamond square algorithm and hydraulic erosion algorithm, an element below for noise generation and to the right a method to save the file in the location of the executable. The UI was placed in the top left as that is the preferred side for important UI for right-handed users which makes up a larger portion of the population along with being a standard in software practices.

At the end of development, the tool was distributed to a variety of participants along with a survey, see appendix 1, there was not a set task assigned for this sprint.

Figure 3: Image of the first version of the terrain generation tool

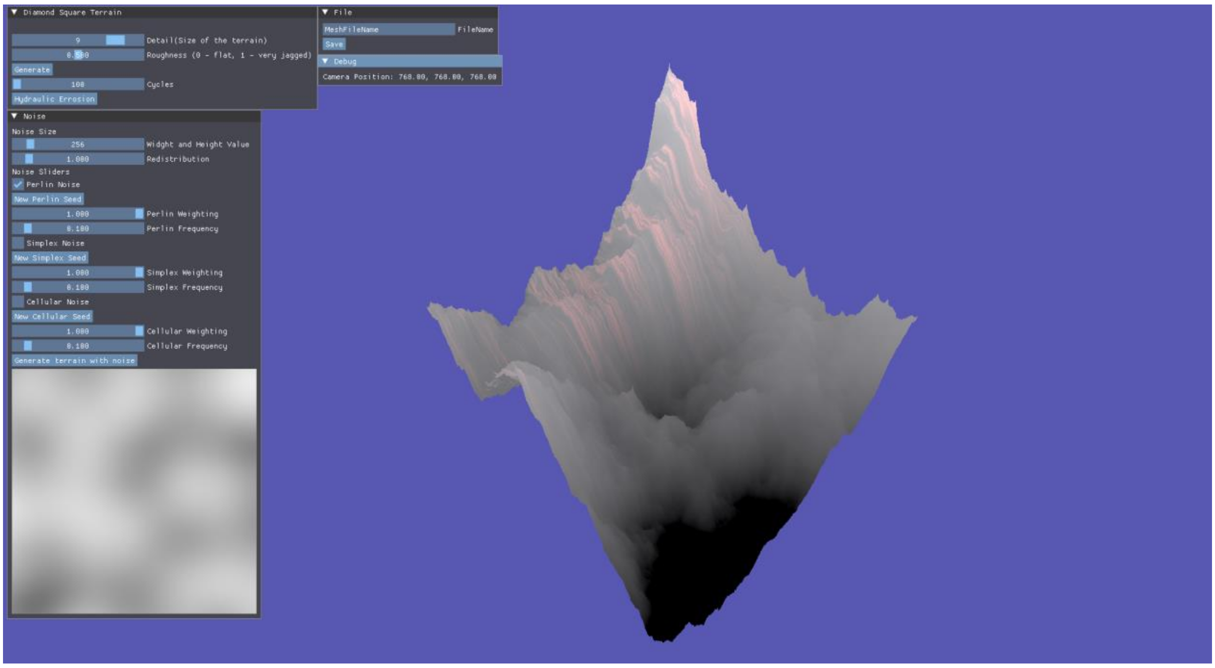


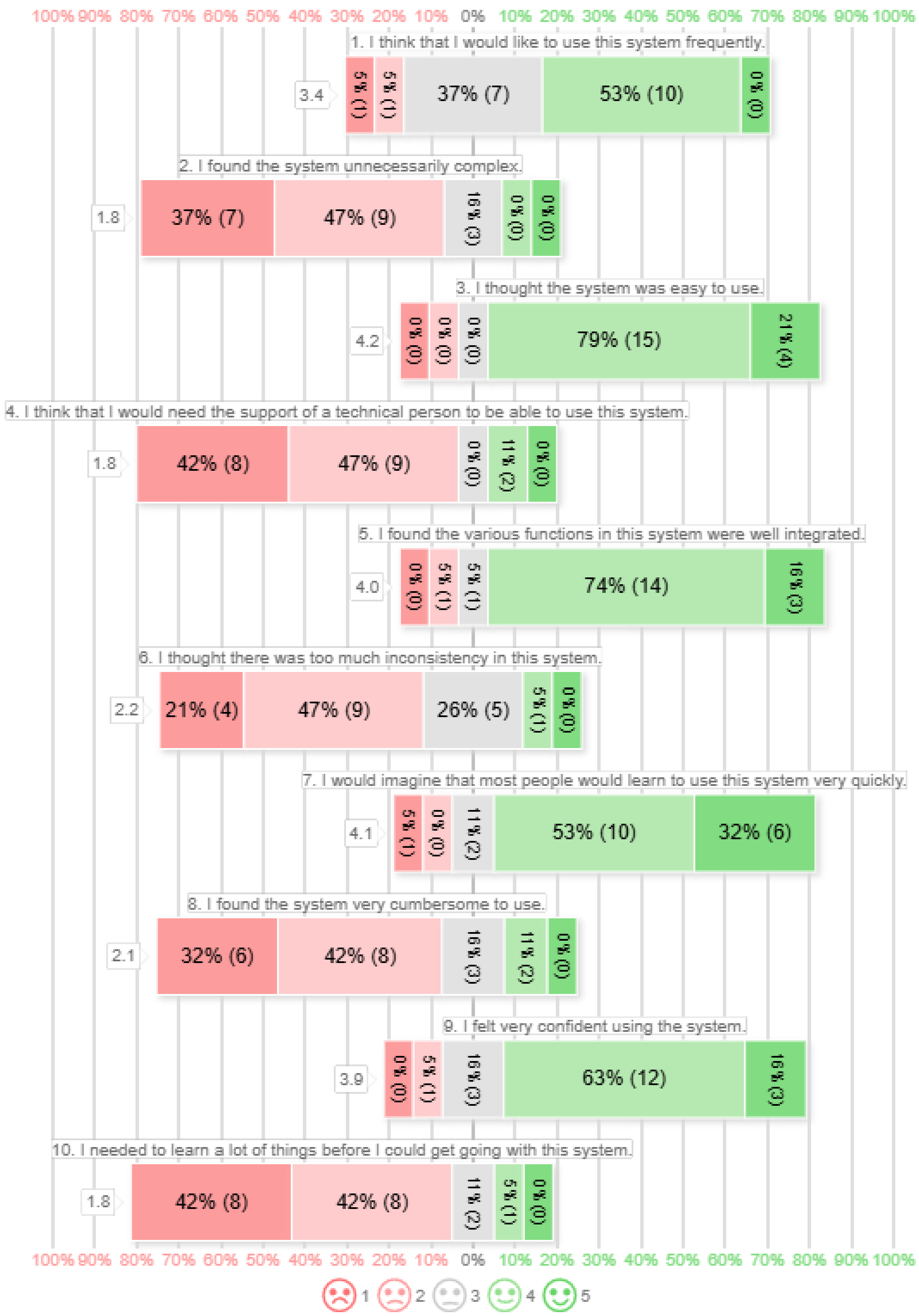
Table 1: Backlog for Sprint 1

Backlog	Sprint 1
Value	Implementation
1	Implement DirectX11 framework
1	Show terrain to screen
1	Diamond Square Algorithm
1	Noise Terrain Generation
1	Hydraulic Erosion
1	Export as OBJ file
1	Basic Camera Movement

Table 2: Table of SUS results for sprint 1

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	4	1	4	2	4	1	5	1	4	2	85
2	3	1	5	1	4	2	4	1	4	1	85
3	4	3	4	2	4	3	3	2	4	3	65
4	3	1	4	1	4	2	1	4	4	1	67.5
5	4	1	5	1	5	2	5	1	5	4	87.5
6	4	3	4	2	4	2	3	3	4	3	65
7	3	2	4	4	3	3	5	2	4	1	67.5
9	4	2	4	1	5	1	5	1	5	1	92.5
11	3	2	4	2	4	2	5	3	4	2	72.5
12	3	2	4	1	4	3	4	2	3	1	72.5
13	3	2	4	2	4	2	4	2	3	2	70
14	4	1	4	1	4	2	4	2	4	1	82.5
15	3	1	5	1	4	3	4	1	5	1	85
16	1	2	4	2	4	2	5	2	4	2	70
17	4	3	5	1	2	4	4	4	4	2	62.5
18	4	1	4	2	4	1	4	1	4	1	85
19	4	2	4	4	5	1	4	2	2	2	70
20	2	2	4	2	4	2	4	2	4	2	70
21	4	2	4	2	4	3	4	3	3	2	67.5
Average	3.37	1.79	4.21	1.79	4.00	2.16	4.05	2.05	3.89	1.79	74.87

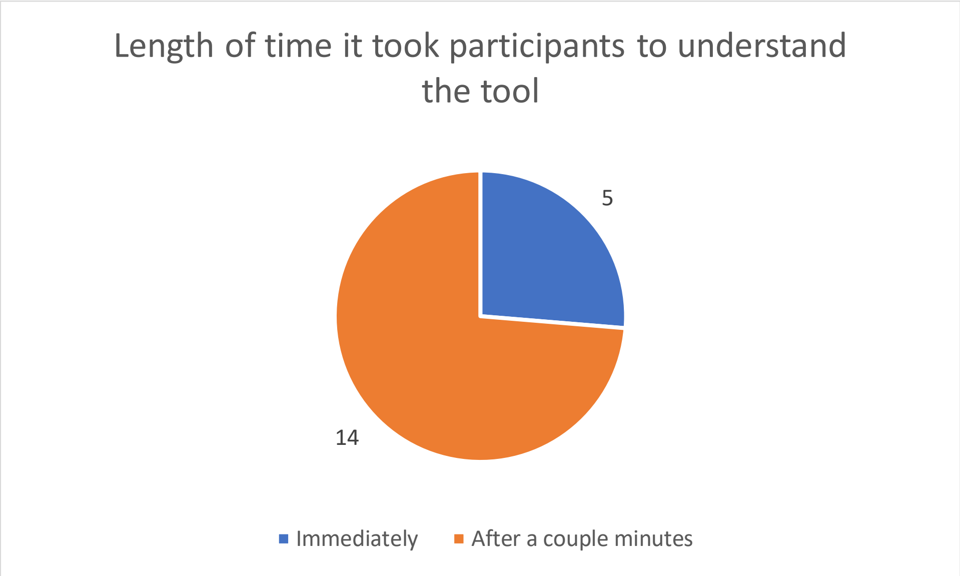
Figure 4: Spread of SUS responses for sprint 1



The resulting SUS score can be placed on the SUS scale shown in Figure 1 to evaluate the performance, Bangor's (2009) scale shows a clear range for the determination of a tool's usability, ranging from A - Best Imaginable to F - Worst Imaginable. A point for consideration is that F contains 3 adjective rankings compared to the top three grades containing singular ratings. The results of this round of testing as can be seen in the Table 2 had a mean of 74.87 which is a C. According to the adjective SUS scale this result can be considered “good” on the Adjective Ranking whereby set score ranges are given word-based values.

In response to the question “How would you change the layout to better work for you?”. Participants put forward three main suggestions for development. Firstly, there was not a clear enough distinction between the two terrain algorithms as participants were getting confused in an attempt to utilise the algorithms together. Secondly, the noise sliders being interactable while a specific noise is not enabled lead to ambiguity as to why the participants’ expected outcome was not occurring. Finally, the interface for hydraulic erosion shares a window with the diamond square algorithm which is ambiguous to its functionality and application.

Figure 5: Pie chart showing the length of time participants took to understand the tool



Participants were asked “How long did it take for you to feel like you understood all the functions of the tool?”, 5 participants said Immediately, whilst 14 participants indicated a couple minutes, this suggests the tool is simple enough and straightforward to navigate in its current version.

Features that users found particularly frustrating or did not work as expected were focused predominantly on the camera. Seven users made suggestions with regards to the camera’s controls not being similar to industry standard modelling software, the controls were a WASD+EQ and mouse right click to rotate the camera which is vastly different to the industry standard which is predominantly an orbital control scheme where users can zoom in and out of a centre model along with changing the rotation point. From the responses it can be inferred that innovating layouts, key binds and interactions may be troublesome as users show preference for layouts and interactions that they are familiar with.

One participant reported issues reading the text on the screen, to accommodate this a different font was used. The font chosen was Calibri, which was recommended in the dyslexia friendly style guide from the British Dyslexia Association.

A single bug was reported, upon inspection, it became evident that it was not a tool bug but was rather caused by the user running the tool in a compressed file, this was subsequently addressed in a later sprint, but was highlighted on the website to inform participants with clear instructions on the correct use of the tool. Suggestions for new features were plentiful, the program’s usability testing is an ongoing cycle rather than a singular event, the insight gained from this round of testing is directly passed onto the next sprints backlog.

Table 3: Backlog for Sprint 2

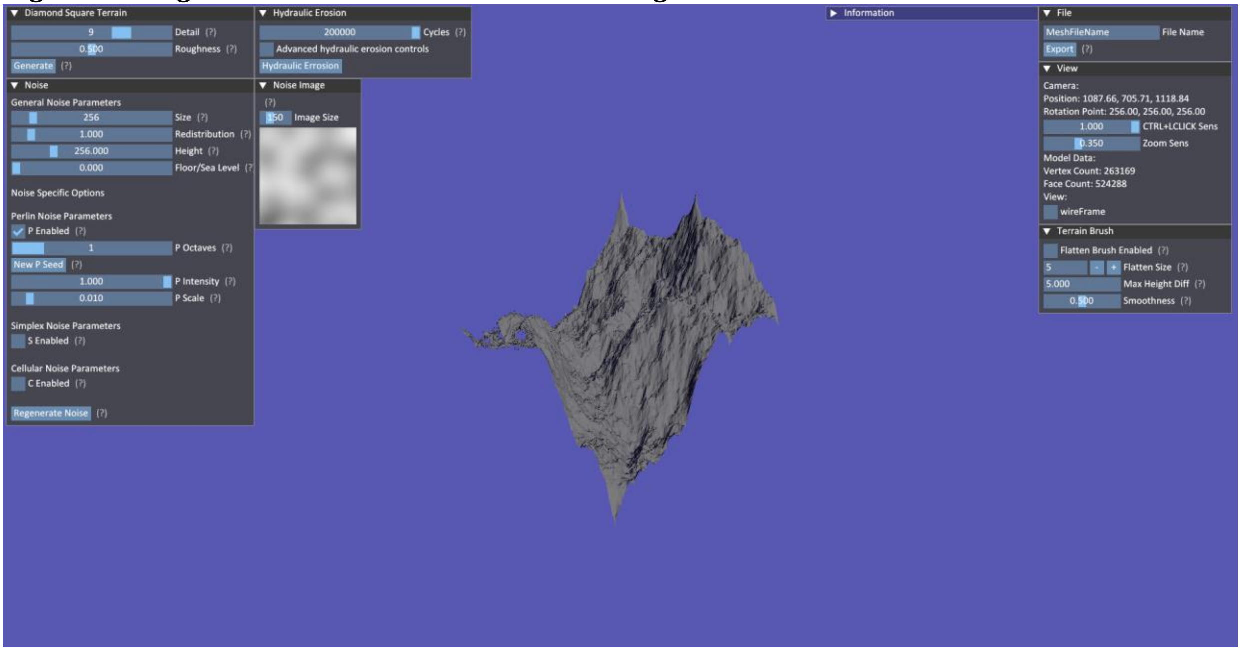
Backlog	Sprint 2
1	Change Camera movement from directional to rotational
1	separate hydraulic erosion from diamond square terrain
1	Add more noise controls
1	Noise UI not shown when not in use
1	improve terrain view
1	Topology view
1	explanations for sliders (tooltips)
2	Implement Brushes
2	Move file save element to the right
2	Different file types
2	Fix lag
2	Clearer Wording
3	Undo/Redo

Sprint 2

After creating the new backlog using feedback from the previous survey, development started. The predominant focus of sprint two addressed user feedback. The most significant changes were to the camera controls along with addressing the lack of information regarding the features of the tool. These changes were considered a priority as usability is not just about functionality, but also about how clearly users understand the tools capabilities. An orbiting camera was implemented to address the cameras usability issues along with a tooltip system which was added to each slider and interactive piece of UI to address the ambiguity in the function of UI elements. Alongside this, a drop-down information tab was added with a button to set specific values to showcase the capabilities of the noise generation system to the participant.

Some of the requested features from the previous survey which were added to the back log, shown in Table 3 such as the undo/redo and full brush implementation were pushed back due to their complexity allowing for focus on usability improvements foremost. This aligns with the core principes of usability to focus development on improving user experience, prioritising development of the experience over the implementation of new features. Features which were not implemented or were partially implemented were added to the backlog for the next sprint. At the end of the development, the tool was distributed to participants along with a survey, see appendix 2.

Figure 6: Image of the second version of the terrain generation tool



Instructions for task one:

- Use noise, height and redistribution to recreate the terrain in the following image (Figure 7)

Figure 7: Image accompanying task 1 instructions

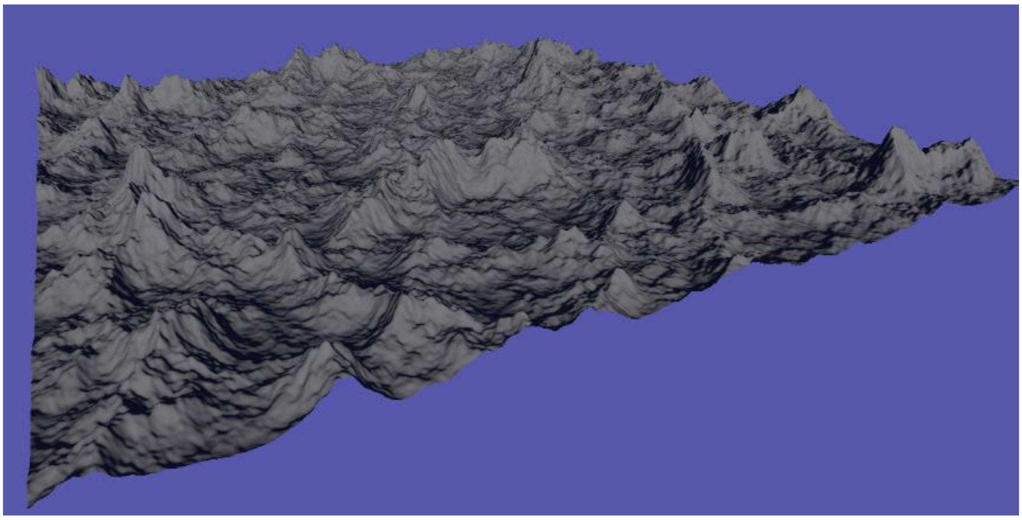
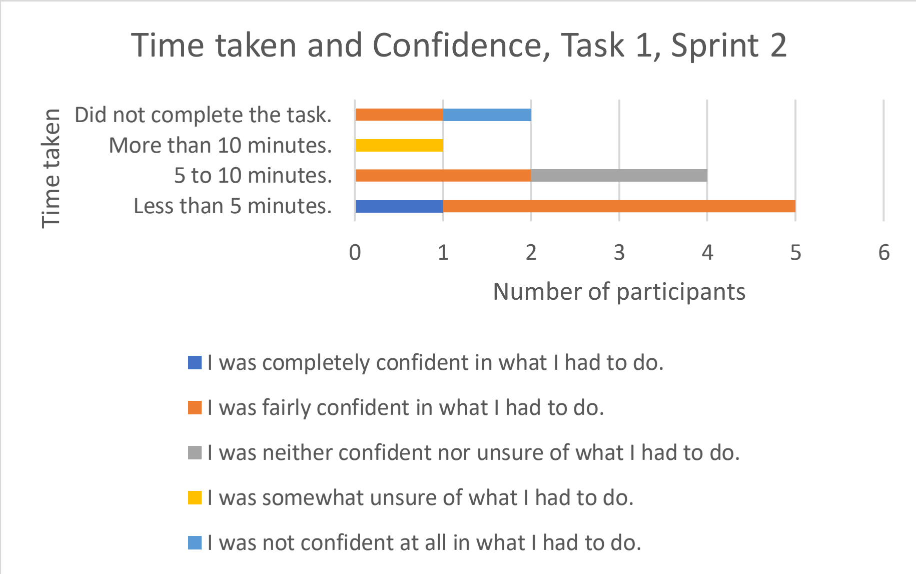


Figure 8: Users confidence compared to time taken in Task 1, Sprint 2

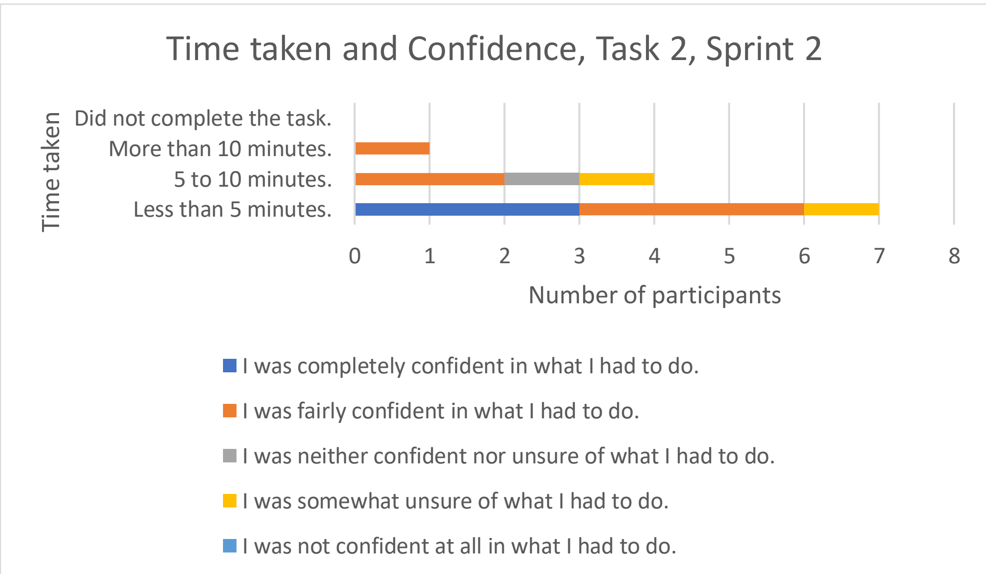


This task was not successfully completed by two users making up 16% of participants, revealing a lack of clarity in the tasks goal. A large portion of the participants still managed to complete the task. The majority of the participants who were unsure, managed to navigate and complete the task. The previous sprint's focus on tooltips and visual feedback may have enabled these participants to complete the task on their own, but suggests that further improvements are needed, particularly in communicating how the noise algorithm's function.

Instruction for task two:

- Create a large terrain with the diamond square algorithm,
- Use hydraulic erosion to make it more realistic,
- Use the flatten brush to create a flat platform on the side of a cliff face

Figure 9: Users confidence compared to time taken in Task 2, Sprint 2

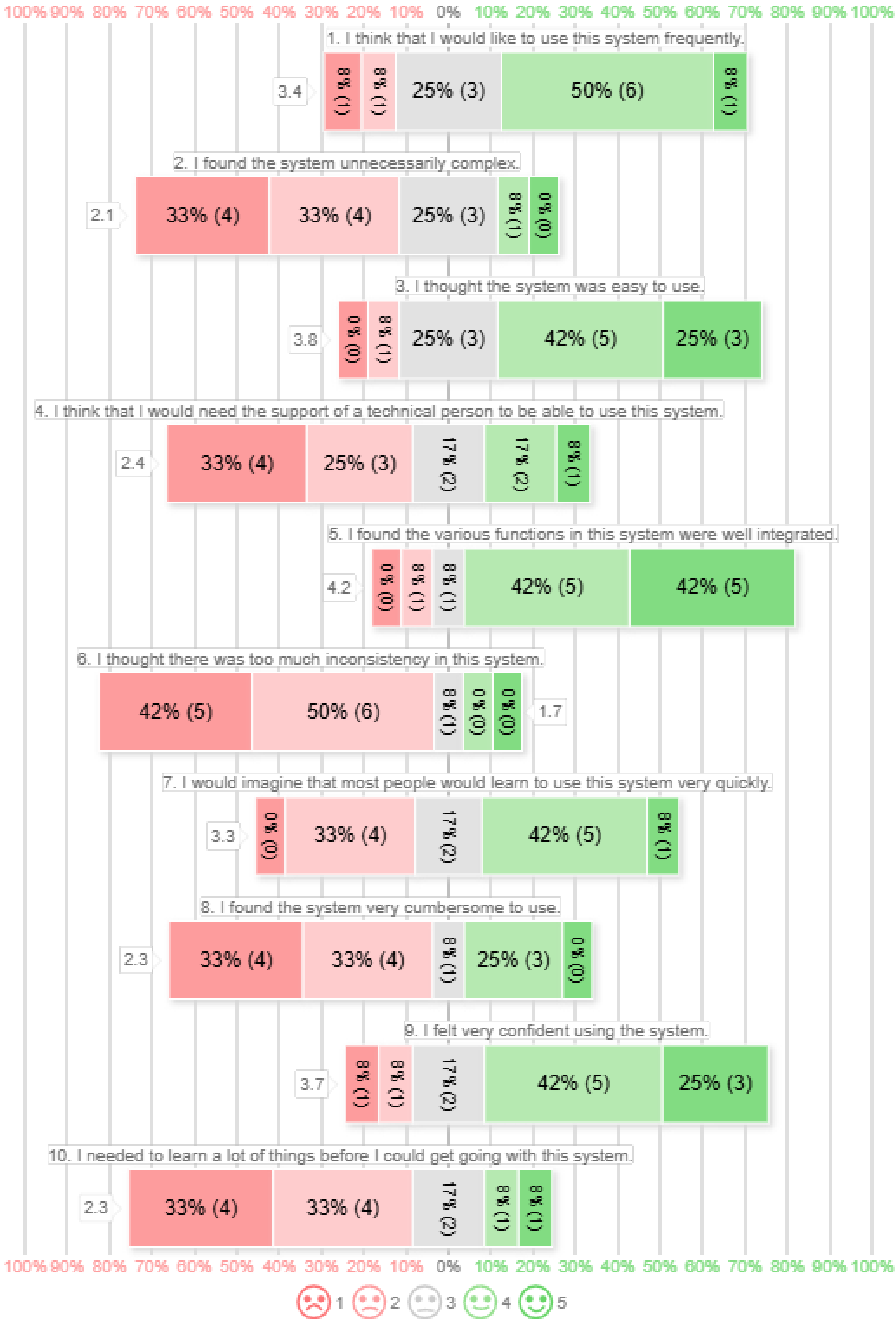


This task had a higher confidence than the previous task as can be seen in figure 9, however the task given is more explicit in steps. This confidence level was reflected in the performance of the task as the majority of participants completed it in under 5 minutes.

Table 4: Table of SUS results for sprint 2

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	4	3	4	4	4	2	2	2	4	4	57.5
2	1	3	2	5	3	2	2	4	1	5	25
3	4	1	4	2	5	2	4	1	4	2	82.5
4	4	1	4	2	5	1	4	1	5	1	90
5	3	2	5	2	4	2	4	1	3	2	75
6	4	2	4	1	5	1	4	2	4	2	82.5
7	4	3	3	4	4	3	3	2	2	1	57.5
8	5	1	5	1	5	1	5	1	5	1	100
9	4	1	5	1	5	1	4	2	5	1	92.5
10	2	4	3	1	4	1	2	4	4	3	55
11	3	2	4	3	2	2	2	3	4	3	55
12	3	2	3	3	4	2	3	4	3	2	57.5
Mean	3.42	2.08	3.83	2.42	4.17	1.67	3.25	2.25	3.67	2.25	69.17

Figure 10: Spread of SUS responses for sprint 2



The mean SUS score for this round of testing was lower than the previous survey at 69.17 as can be seen in Table 4 which is 5.70 less then the sprint one average of 74.87 as seen in Table 2, this could be due to participants deeper usage of the tool encouraged by the tasks. The standard deviation for the SUS score in this sprint was 20.57. Interestingly, the SUS answers were more distributed as shown in Figure 10 this visual representation of the spread of SUS answers shows a more split opinion by participants than the previous survey.

Figure 11: Users responses comparing this version usability and functionality to the previous versions

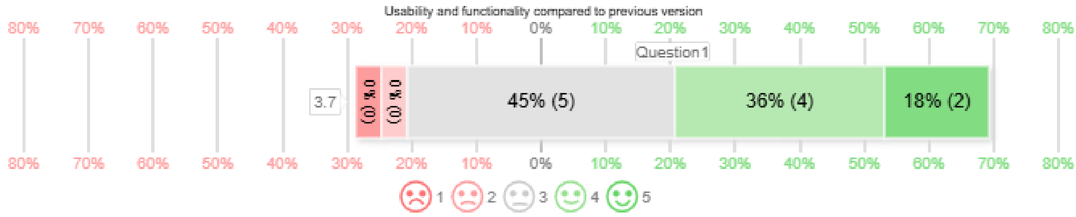
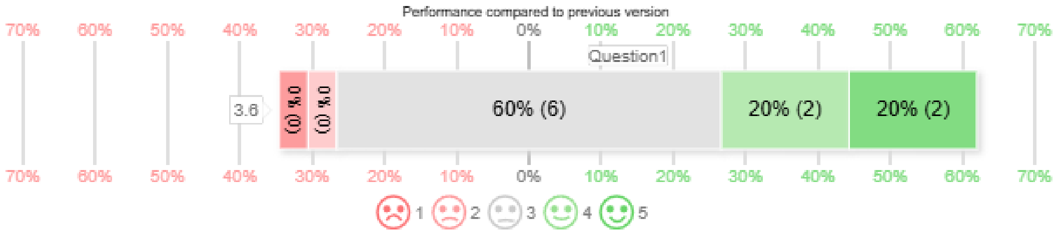


Figure 12: Users responses comparing this version performance to the previous versions



In comparison to the previous study 45.5% of participants said that the usability and functionality of the tool has remained the same whilst 36.4% of participants said there was some improvement and the remaining 18.2% of participants said it was much better. Performance of the tool compared to the previous version has 60% saying the same whilst there's a 20% on both slight and much better.

The participants perceived time to understand the tool was generally similar with an outlier saying that they did not understand the tool.

Layout adjustments suggested by participants: relocate the information box to the top left, all terrain related tools to be near or integrated into the same window, naming conventions simpler and develop the export window.

Additional performance issues that were mentioned were related to the hydraulic erosion taking 10 – 15 seconds, along with when large maps (greater than 1024x1024) were generated performance would drop substantially.

Frustrations with the tool were partially due to a lack of clarification and participant misunderstanding. Due to the limitations of the current application, use of the tool on a secondary monitor makes the camera jump every time you try to rotate it, in addition a concern was mentioned where the UI was incorrectly scaled on 21:9 monitors.

In response to the question asking for suggestions, participants requested: additional brushes along with basic shapes and geological features such as mountains and rivers, a tutorial, Undo/Redo, colour and texture terrain features, focus and centred camera rotation, file loading and additional file types. Which were added to the backlog seen in Table 5.

Table 5: Backlog for sprint 3

Backlog	Sprint 3
2	Undo/Redo
2	Fix lag
2	Different file types
1	Implement Brushes
1	Additional camera controls: focus and centre the camera
3	Add geological features
1	Create Tutorial
2	Colour the terrain
1	File loading

Sprint 3

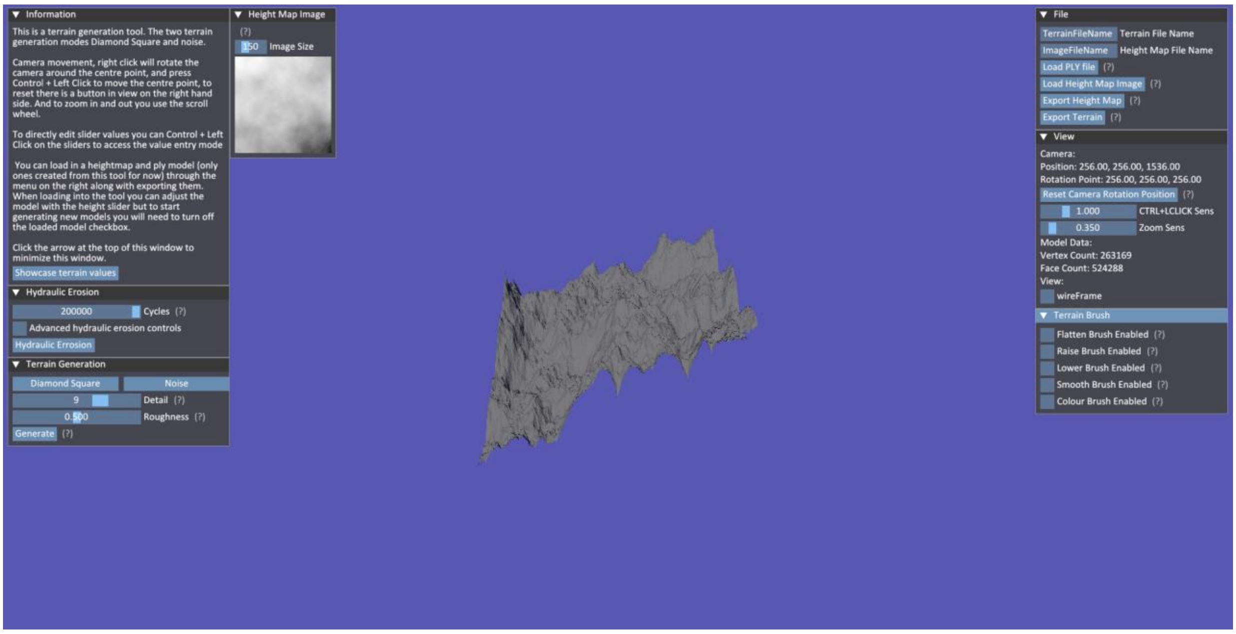
After creating the backlog using feedback from the previous survey, development started. The development of the tool's usability continues by further iterating the UI, development of the exporting and importing feature and the completion of the terrain brushes.

UI changes were made in order to condense the windows along with eliminating the possibility of confusion between the two terrain generation features, which was prevalent in sprint 1. This was resolved by combing the two terrain generation algorithms into a single window with tabs to choose between the two algorithms. Information was moved to the top left side of the screen as requested by participants along with more elaborate descriptions on controls such as the camera and importing and exporting features.

For the import and export system, a library was utilised to allow for the tool to interface with the windows file explorer allowing users to dynamically select both the file to load in and the export location. In the current iteration, entering the files name through the file explorer was not implemented. This enabled users to export the height map as a JPEG and the terrain as a PLY file, along with importing a height map and other height map images into the tool and importing the exported PLY file back into the tool.

In addition, brush features were expanded upon with the smooth, raise, lower and colour brush. These features were placed in the bottom right along with the original flatten tool developed in the previous sprint. This marks the end of the implementation of new terrain editing features as the subsequent sprint will focus on polishing the tool for its for its final iteration. At the end of this round of development, the tool was distributed to participants along with a survey, see appendix 3.

Figure 13: Image of the third version of the terrain generation tool



Instructions for task one:

- Generate a diamond square terrain,
- Use hydraulic erosion to erode the terrain,
- Use the smooth brush to smooth the terrain as hydraulic erosion is quite bumpy (you can use a high brush size),
- Then colour in what you would like to be a river,
- Finally export it as a ply file and then load the file you just saved back into the tool

Figure 13: Image showing an example for the end product for task 1, sprint 3

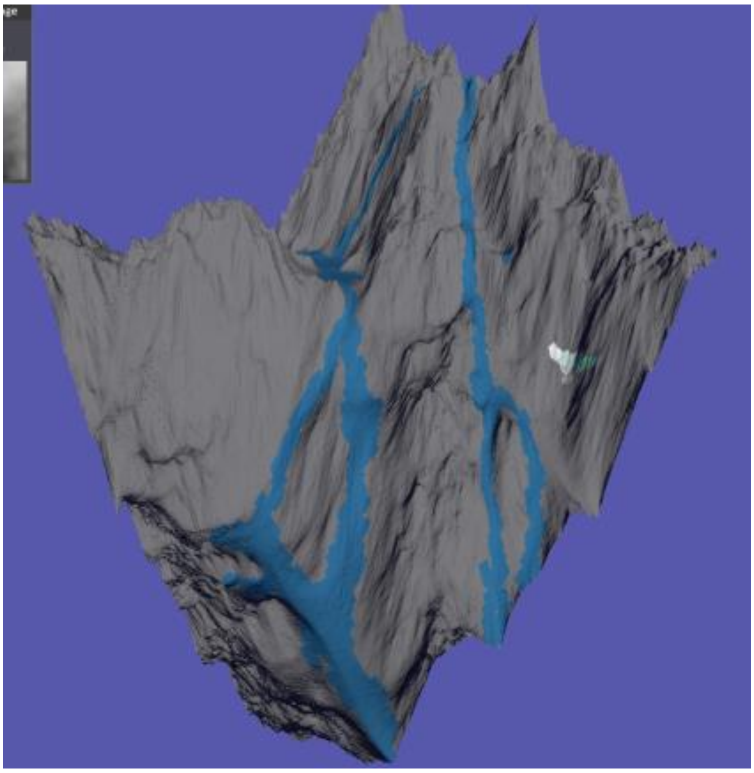
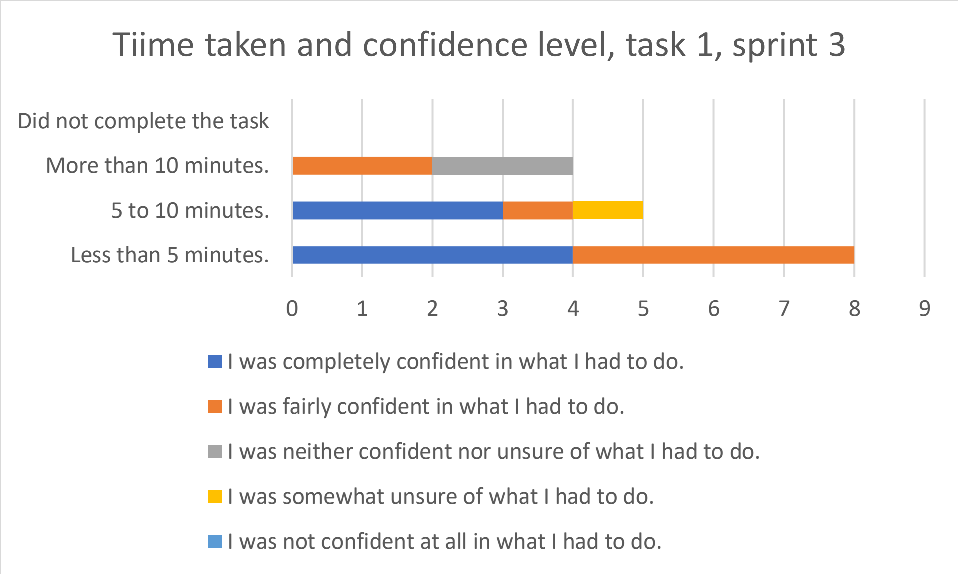


Figure 15: Users confidence compared to time taken in task 1, sprint 3



Task one focused on the use of diamond square algorithm as the base of the task and participants were then guided to use the rest of the tools features to mould the terrain into a more finished product.

Task two:

- Generate a noise terrain that is reasonably mountainous,
- Then use hydraulic erosion once more,
- Followed by that use the raise brush to connect some of the mountains and use the lower brush to create some valleys,
- Save this as a height map and reload the height map back into the tool,
- Adjust the noise with the height slider to look as close to the original terrain as possible

Figure 16: Image showing an example of the terrain after step 1 of task 2, sprint 3

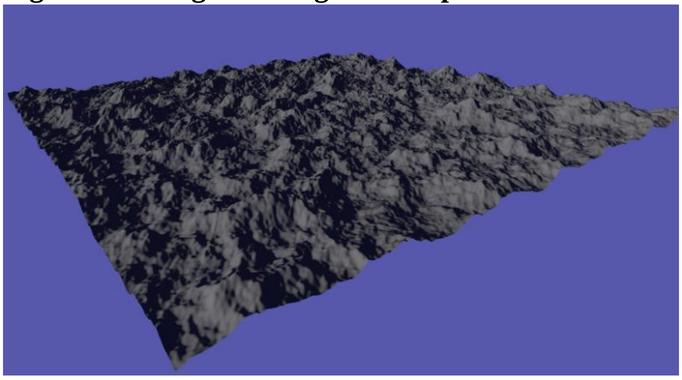
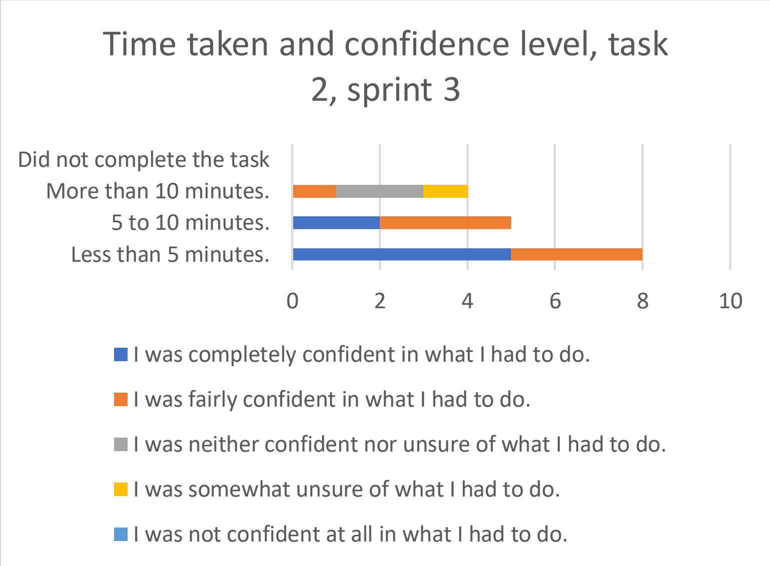


Figure 17: Users confidence compared to time taken in task 2, sprint 3

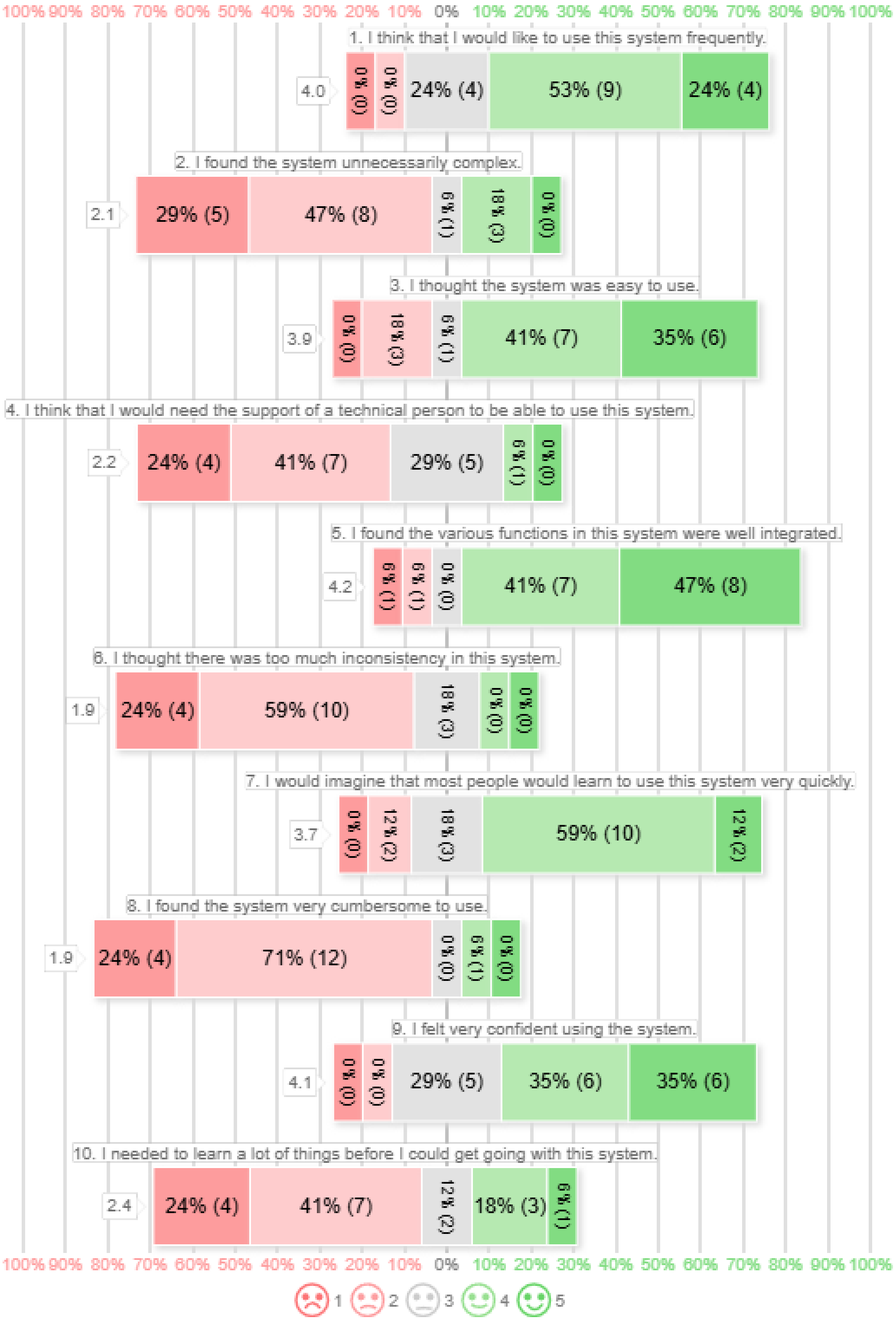


The results for task two are similar to that of task one as the share of participants in each time barrier is the same as shown on Figure 15 and Figure 17.

<u>Table 6: Table of SUS results for sprint 3</u>

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	5	1	5	1	4	1	4	2	4	2	87.5
2	4	2	4	3	4	2	3	2	4	2	70
3	4	2	2	2	5	2	4	2	3	2	70
4	4	1	5	2	4	2	3	2	4	4	72.5
5	5	2	5	2	5	2	4	2	5	2	85
6	4	2	4	3	5	2	4	2	4	2	75
7	4	4	3	2	4	2	3	2	3	4	57.5
8	5	1	4	2	5	1	4	1	5	5	82.5
9	4	2	5	4	5	2	5	1	5	1	85
10	3	2	4	2	4	2	4	2	3	2	70
11	3	4	2	3	1	2	4	4	3	3	42.5
12	4	1	5	1	4	3	4	2	5	1	85
13	3	2	4	2	4	3	2	2	4	3	62.5
14	4	3	4	3	5	2	4	2	4	2	72.5
15	5	2	5	1	5	1	5	1	5	4	90
16	3	4	2	3	2	3	2	2	3	1	47.5
17	4	1	4	1	5	1	4	1	5	1	92.5
Mean	4.00	2.12	3.94	2.18	4.18	1.94	3.71	1.88	4.06	2.41	73.38

Figure 18: Spread of SUS responses for sprint 3



There was a substantial increase to the SUS score of 4.22, it was within the adjective range of ‘good’ and is 1.49 points below the score from original survey. We can see the spread of data in Figure 18 for the SUS has a lot less spread between both the negative and positive choices and lean heavily to the preferred answers with a high correlation to the first questionnaire.

Upon evaluation of the responses, it supports the focus on the tool's usability. In the majority of data relating to changes of the UI, participants communicated that they were satisfied with the layout and that changes were made to accommodate their suggestions. It is notable that participants find it rewarding to see their suggestions implemented through development of the tool.

Frustrations given by participants were significantly lower than in previous sprints and fewer items were mentioned and were accompanied with appreciation. A participant stated that they appreciated the colour brush but identified the lack of an undo button which made correcting mistakes more difficult as they needed to estimate the colour needed. The participant suggested for this was a colour picker tool. One participant suggested that the controls be WASD to move instead of the orbital camera, this was considered by adding an option to allow the user to choose their camera setting. This was eliminated as the majority of participants usability preferences were priority and with the added complexity that accompanies additional camera controls it was not viable. Furthermore, users requested the integration of the file name into the saving process instead of a separate text input element. Finally, one participant mentioned the colour brush was drawing as a square instead of a circle which was not intended and was added to the backlog.

There were no bugs reported that were directly caused by the tool but there were two reports of crashes along with a known issue of the inconsistencies with using the tool on a secondary monitor, this issue was brought to the participants attention on the website.

One user suggested tooltips which was already implemented, throughout development it was noticed that participants consistently requested features that were already implemented into the tool, we can take this information and assume that the features are not clear enough or intuitive enough for users who may not know ahead of time what a specific feature is. Suggestions related to controls were that terrain brushes felt unwieldly, shortcuts would streamline the use of editing features, visual indication of the size of the brushes would make it easier to use. Three features were requested, colour by height value, paint in rivers and lakes, and Undo and Redo. All suggestions were consolidated onto backlog 4 which can be seen below in Table 7.

Table 7: Backlog for sprint 4

Backlog	Sprint 4
1	File name in file explorer
1	Fix Lag
1	Round the colour brush
2	Colour terrain by height
3	Placeable terrain features
1	Make the brushes values feel better
1	Undo Redo
1	Key Bindings
1	Smooth all terrain button
1	Move the brush UI elements to the top

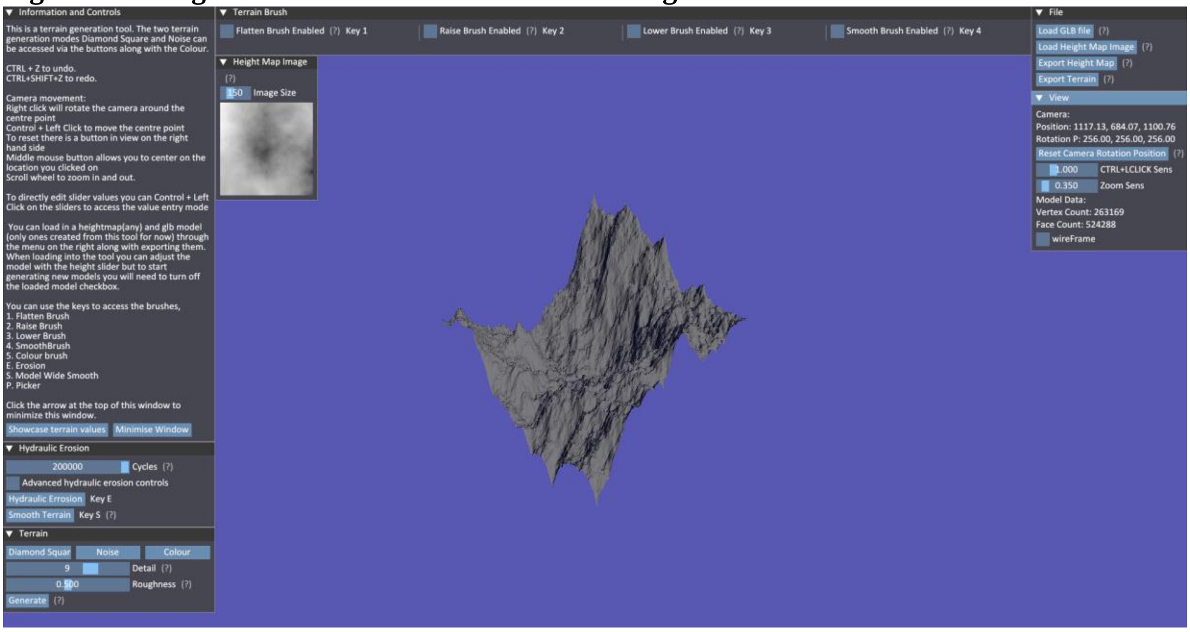
Sprint 4

The focus for the final round of testing, was the consolidation and resolution of usability issues along with reviewing features with existing problems and rounding up the backlog. Two features that users requested for enhancement for an extended period were a performance fix for the noise sliders as well as an undo/redo function, which were both subsequently added. The slider solution scales down the terrain while the slider is held which improves performance, when the slider is let go it generates the full resolution terrain.

Other usability changes incorporated the addition of shortcuts along with the information section which was expanded to contain more information along with all the controls such as the shortcuts, undo/redo and camera movement.

The file saving function was developed to align more closely with industry standards through the removal of the file name text input and integrating it with windows file explorer. At the end of the final round of development, the tool was distributed to participants along with a survey, see appendix 4.

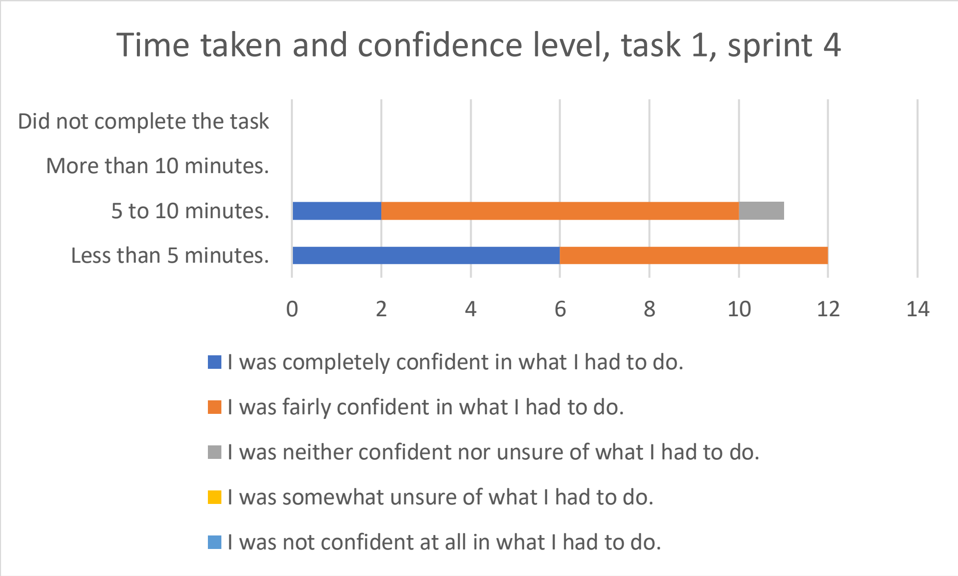
Figure 19: Image of the fourth version of the terrain generation tool



Task one:

- Generate a diamond square terrain,
- Use the hydraulic erosion key shortcut,
- Use the smooth all terrain shortcut to smooth the terrain out,
- Manipulate the terrain with the brushes using the shortcuts,
- Undo and Redo some of your edits,
- Export it,
- Generate a new terrain,
- Load your exported terrain back into the tool,
- Manipulate it with the height slider

Figure 17: Users confidence compared to time taken in Task 1, Sprint 4

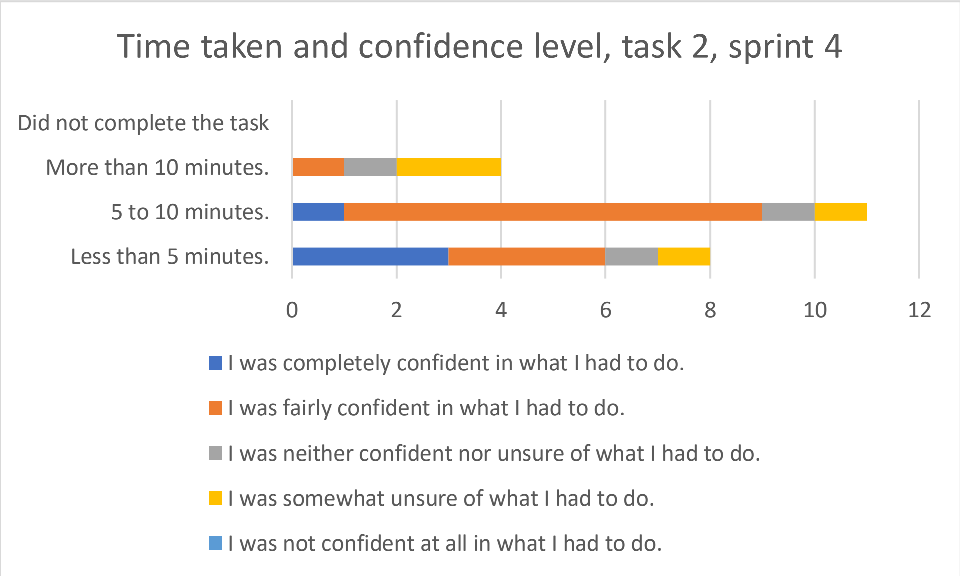


Task one's performance or time taken was a lot more condensed for this round then previously even though the tasks have gotten longer and more complex.

Task two:

- Generate a large noise terrain using two or more noises with octaves,
- Adjust this terrain to have mountains and flat planes,
- Create a river system and lakes,
- Set the colour of the terrain,
- Export the heightmap,
- Load it back in,
- Adjust it with the height slider to replicate how it was previously

Figure 21: Confidence level of participants in Task 2, Sprint 4

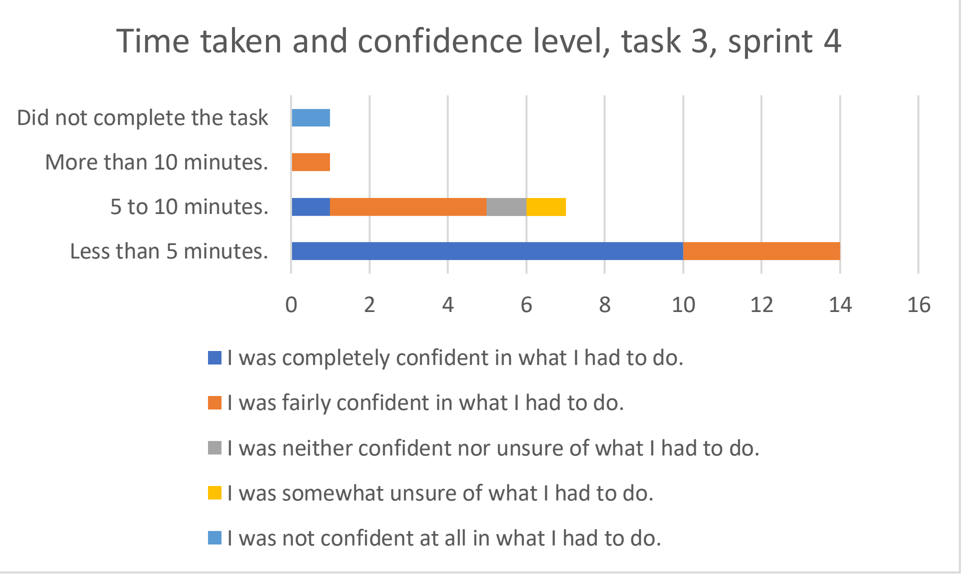


Task two, includes the use of noise, this use within the task causes the time taken to increase and vary quite a lot compared to the diamond square algorithm, this is reflected in the previous results for tasks using noise.

Task 3:

- Find a (square)heightmap image online,
- Adjust the height

Figure 22: Confidence level of participants in Task 3, Sprint 4

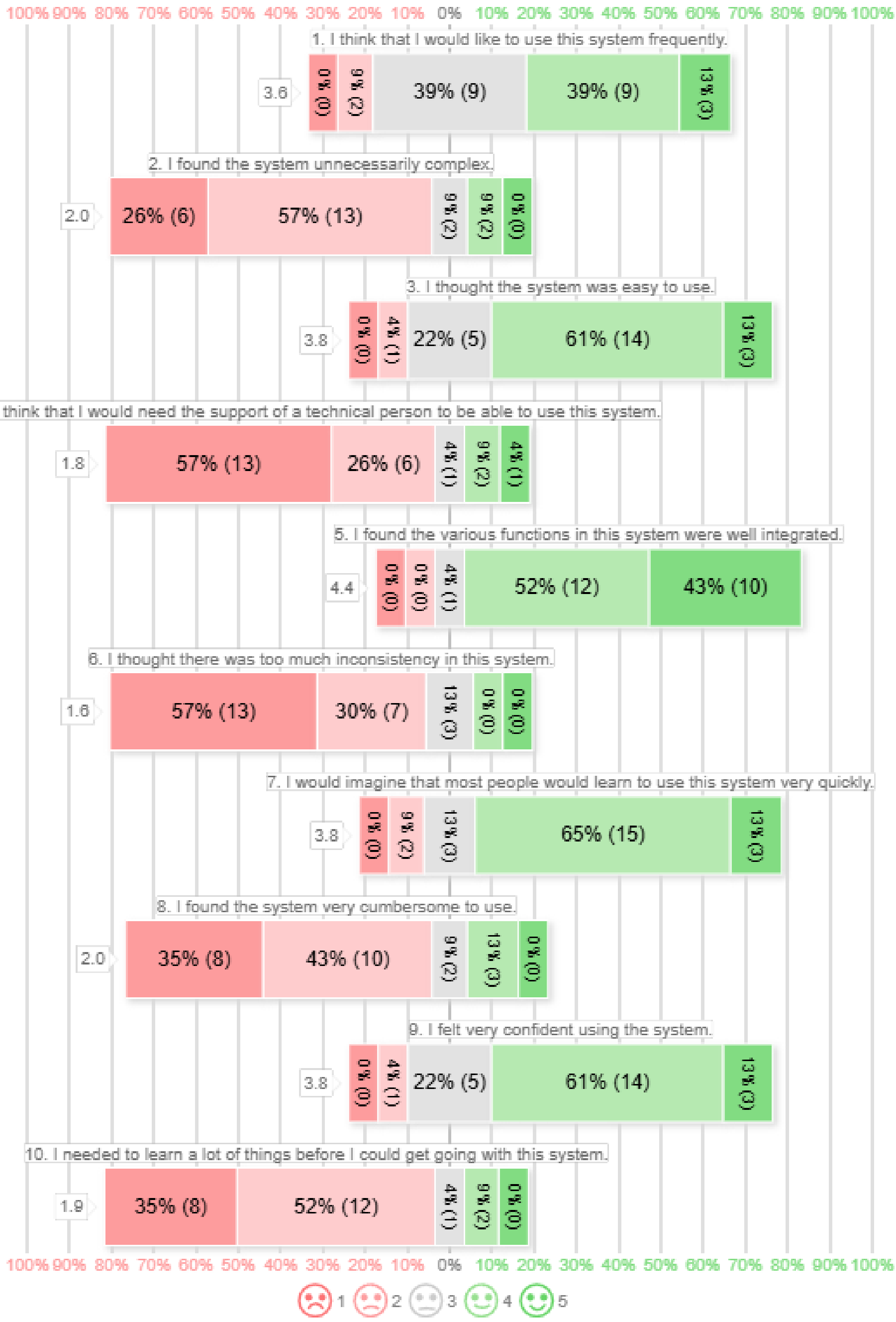


We can see the completion time for Task 3, Sprint 4 is mostly less than 5 minutes which may be a result of the simplicity of the task. Accompanying this simplicity, an issue was identified with the task, that is the task itself assumes participants know what a heightmap is which could have been explained in greater detail leading to the possible improvements in confidence of users in the task.

Table 8: Table of SUS results for sprint 4

Round 4										
ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
1	4	2	4	1	5	1	4	2	4	1	85
2	4	2	4	2	4	2	4	2	4	2	75
3	4	1	4	1	5	1	4	1	5	2	90
4	4	1	5	1	5	1	5	2	4	2	90
5	4	1	5	1	5	1	4	1	4	1	92.5
6	5	2	4	2	4	1	4	1	4	2	82.5
7	5	2	4	1	4	1	4	1	5	1	90
8	3	3	3	3	4	2	4	3	4	3	60
9	5	3	4	4	4	2	5	4	4	4	62.5
10	4	4	3	1	5	1	3	2	4	2	72.5
11	2	2	3	2	4	1	4	1	4	1	75
12	4	2	4	1	5	3	4	2	4	2	77.5
13	4	2	4	1	5	2	4	2	4	1	82.5
14	2	2	4	5	4	2	3	2	3	2	57.5
15	3	2	4	1	4	1	4	4	3	2	70
16	3	4	2	2	4	3	2	4	2	2	45
17	4	1	3	1	5	1	3	2	3	2	77.5
18	3	2	4	2	4	1	4	1	4	1	80
19	3	2	4	1	4	1	4	2	4	4	72.5
20	3	2	4	2	3	3	4	2	4	2	67.5
21	3	1	3	1	4	2	4	1	3	1	77.5
22	3	2	4	4	5	2	2	3	3	2	60
23	3	1	5	1	5	1	5	1	5	1	95
Mean	3.57	2.00	3.83	1.78	4.39	1.57	3.83	2.00	3.83	1.87	75.54

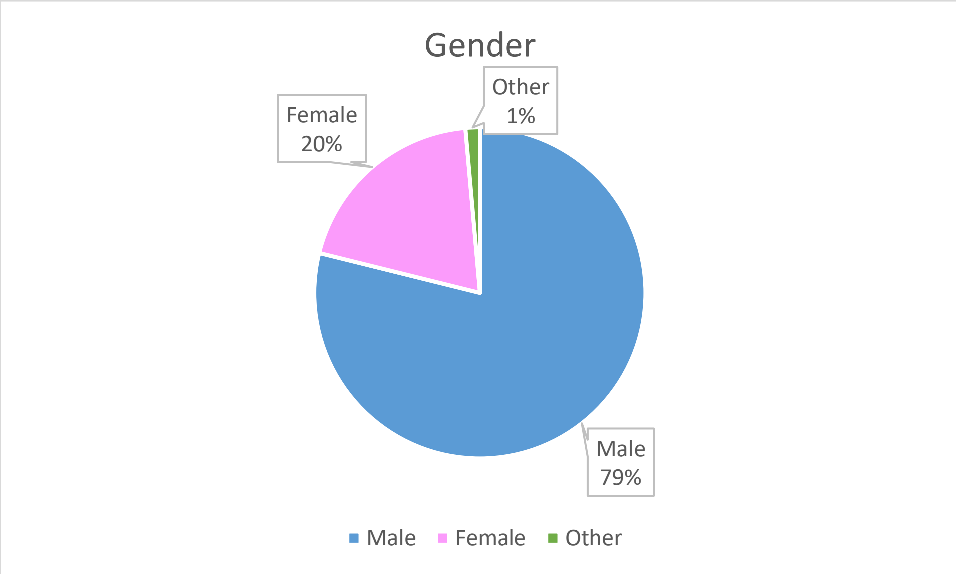
Figure 23: Spread of SUS responses for sprint 4



This sprints SUS questionnaire results are an improvement of the last and have overtaken the high score of 74.89 from sprint 1 with 75.54. Supporting this increase in SUS score we can see in Figure 23 that user's responses are more condensed in the questions.

After the final sprint, the two features that weren't implemented were as follows: Colour by height, allowing users to set height boundaries where those colours would be set, and placeable terrain features, these weren't added due to the complexity of the features along with the time constraints of development timeline.

Figure 24: Pie chart gender split for the whole study



During the duration of the study, there were 73 responses and 71 people who consented to participate and filled the rest of the form out. The age range of participants was between 18 – 64 years of age with a mean of 27.29 along with a standard deviation 11.71 which is a reasonable deviation for age. For all participants the split of gender was 56 male participants, 14 female participants and one other as can be seen in Figure 24. Most participants were students and teachers, with a few outliers such as a Senior Systems Analyst, Export Clerk and Artist.

Discussion and Analysis

Sprint 1

Upon the completion of the first sprint the tool had its core mechanics implemented; this was enough to start the usability testing of the tool. The majority of iterative development studies do usability testing throughout the study but use the SUS questionnaire exclusively in the final round of testing, this study diverged from this and included the SUS questionnaire in each round to allow for analysis and the comparison between the SUS scores.

This sprints SUS score had a high mean score of 74.87 which may have been related to the limited number of features (diamond square terrain generation, noise terrain generation and hydraulic erosion) usable by participant which results in the tools overall simplicity. This raises an interesting dilemma; how do you balance the usability of a tool and the depth of the features? A tool that is too simple would limit the user whilst a tool that is overly complex would deter new users, the goal is to find the middle ground in complexity and usability. This round focused on complexity due to the requirement of meeting the tools minimum required features in comparison to later sprints.

Other notable data includes, the question evaluating the participants time to understand the tool the results of which 5 participants responded 'immediately' and 14 participants responded a 'couple of minutes'. The remaining unselected choices with the other two options being choices were: 'more than 10 minutes and' 'I did not understand the tool'. This indicates the tools overall simplicity and straightforward nature which allowed users to navigate and learn on their own.

With regards to the question, "what participants find frustrating and the changes they want?", a large majority mentioned the camera as an obstacle. Identifying the majority of participants found one aspect frustrating indicates an overhaul for the camera is needed. Due to the majority consensus on the camera, the inference can be made that users who did not report the camera as an obstacle shared the frustration.

One participant reported a bug concerning their malfunctioning terrain exportation feature, through testing this is not a direct fault from the tool and rather a misunderstanding by the user and a gap in the information regarding the use of the tool. From this it can be ascertained that some usability issues are not related to the tool itself but rather how users interact with it which emphasises the need for adequate accessible information to the user about all features and actions.

Sprint 2

Tasks were integrated into this round of testing and onward to encourage deeper use of the tool. Task one was not successfully completed by all participants, Figure 7, revealing a lack of clarity in the tasks goal. The task is ambiguous about whether the participant was to aim for an exact match of the terrain in Figure 6 or just something similar. Despite this ambiguity, a large portion of participants managed to complete the task, which in turn shows a strong learnability of the tool. Many participants who reported themselves as unsure managed to identify the correct steps for completion of the task suggesting the tool supports intuitive learning. It can be deducted, that even a well-designed tool can suffer from unclear instructions. Ensuring tasks are clear is just as important as the clarity of the UI when testing usability. The participants ability to deduce the steps for completion may have been supported by the addition of tooltips in Sprint 2, feedback suggests that further improvements are needed, particularly in the communication on how the noise algorithm's work.

Task two in contrast had a higher number of participants respond that they were more confident, accompanied by accomplishing the task in less time than the previous task as can be seen in Figure 8. It can be derived that the confidence of the participant could be due to the simplicity of the features that participants were asked to use in comparison to task 1 which relied on noise generation which is more complex.

Within this round of testing, we can see that the SUS score had a decline from 74.87 to 69.17. A major factor could be the integration of tasks into the usage of the tool, this guides the user to engage with tool more deeply. In addition, the complexity of the tasks could lead to a higher perceived difficulty. This resulted in users who struggled especially on task one having higher likelihood to view the tool as less usable.

Improvements were noted as 6 out of 11 users responded that the tool was slightly and/or much better than the previous iteration whilst the remaining 5 said it was 'about the same'. This improvement may have increased the participants expectations of the tool leading to smaller issues becoming more evident. As recurring participants engage in the survey their understanding of the tool and its capabilities broaden, they engage with the tool more leading to identification of more niche frustrations that weren't noticeable through their surface level use during the first survey. The preferred outcome would have been for the SUS score to increase but this decline provides scope for further development and to increase the tools usability. It can be inferred that a lower SUS score does not always signify that users struggle, it can reflect deeper use or higher expectations.

Sprint 3

Complexity of the tool has increased after Sprint 3 due the implementation of further features such as the addition of terrain brushes. In response to increased complexity, tasks have mirrored the increase as they accommodate the use of a wider range of features. In response to the findings of the previous round, tasks have been written to be more explicit and straight forward leading to a simpler experience for participants; this is further supported by the ongoing development of the tool's user experience.

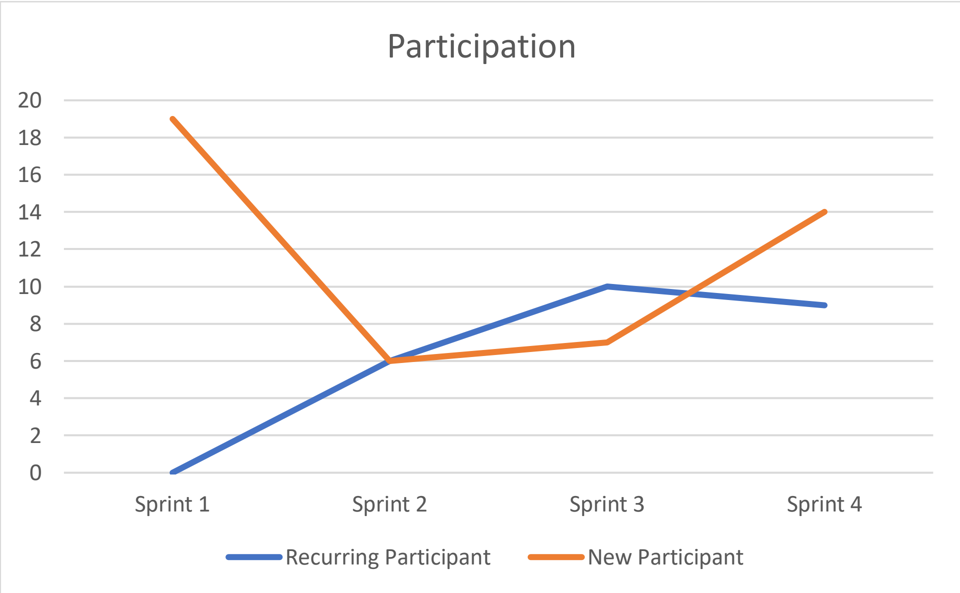
A large portion of participants completed the tasks in under 10 minutes as shown in Figure 15 and Figure 17. It can be observed, one participant stated they were unsure of what they had to do but still completed the task within 5 to 10 minutes. This suggests the tool enables participants to learn through its use and the information provided on features. A point for consideration is that as tasks are both longer and provide participants freedom to determine the time spent on each section, an increase in time spent on tasks cannot be assumed to correlate with the difficulty of participants experiences. If time is to be a metric of usability, constraints should be put on the user to do it as fast as possible supported by the reduction of randomness in the task.

This round of testing indicated an increase in the tool's overall usability as the SUS score increases standing just below the original score at 73.38. This highlights the impact of constant iteration on the usability of the tool as the complexity has increased alongside the task whilst the usability score has improved.

Sprint 4

Sprint 4 being the final round of testing, the emphasis was on polishing the tools interface and functionality. This meant major features were not added in favour of developing usability.

Figure 25: Chart showing the number of new participants and recurring participants through the study



In task one, participants were more confident than previous rounds as can be seen in Figure 20, this improvement could be attributed increased number of recurring participants in comparison to new ones. However, this sprint had only 9 recurring participants and 14 new participants as shown in Figure 25, which indicates that the confidence of participants is not from experience only but rather the tools learnability. The completion rate in conjunction with the confidence are within a smaller range of times, 'less than 5 minutes' and '5 to 10 minutes' compared to tasks from previous sprints. It can be ascertained to show that the workflow of the tool has reached an equilibrium in which the users experience does not heavily influence the time of completion but instead the tool and the task itself, which supports that the tool is both streamlined and learnable to all users.

Task two has a more varied time taken, <5, 5-10, and >10 compared to task ones <5 and 5-10, which could be attributed to the complexity of noise generation along with the freedom available to the participant. As mentioned previously it cannot be assumed that an increased time directly correlates with frustration or struggle with features contained in the task. Instead, a question asking whether the user felt frustrated completing the task along with where in the task that may have taken place could benefit the evaluation.

Task three in retrospect relies on the users understanding of what a heightmap is, this assumption of knowledge means the task itself is suboptimal as there are expectations placed on participants. This could be improved by introducing participants to concepts which can be explained and shown through visual examples and descriptions. This would minimise the reliance on prior knowledge and leave all participants on an equal playing field which may result in more consistent usability results.

The SUS score for the final round of testing had a mean of 75.54 which is 0.67 above the original score of 74.87, this outcome is indicative of the iteration and focus on usability done throughout the study. With the overall increase of the complexity of the tool, increased usability score and large population of new participants, it can be concluded that the overall effectiveness of usability focused development has shown positive results, demonstrating its significant impact on improving user satisfaction, efficiency and usability.

Conclusion

This study set out to improve the usability of a terrain generation tool through iterative development and user feedback. Through the collection of user feedback at each stage, the tool improved from a simple tool with some algorithms to a more developed and complete tool whilst maintaining and improving the usability. Incorporating SUS scores and structured tasks provided measurable data on how participants interacted with the tool highlighting the challenge of balancing depth and accessibility.

A key finding is that usability is not solely dependent on one factor. The first SUS score was high due to the tool's overall simplicity, this score dropped after new features were implemented highlighting the challenge of maintaining usability while increasing functionality. From this we can gather that adding complexity doesn't have to sacrifice the tools usability as long as the users are properly informed of the tool's functionality. The final SUS score being higher than the original shows that an iterative approach to usability can work, if user feedback is consistently acted upon.

Upon evaluating task data, longer completion time do not mean users were struggling, it could be due to the nature of the task, or the freedom given to the user. A big factor was that if a task wasn't clear, users took longer, not exclusively because of the tool's complexity but because of how the task was described. This reinforces the idea that usability isn't just about how the tool is built but also how it is explained to users. With this in mind, this dissertation has satisfied identification of core aspects that affect the usability of a terrain generation tool.

This study adds to the growing body of research and has indicated that focusing on usability throughout the development process makes a difference to the terrain generation tool's overall usability and learnability which aligns with the findings from Blomqvistand and Detterfelt (2020) focusing on the production of a game development tool through usability testing. By listening to users and making changes based on their experience, a terrain generation tool can become more complex without sacrificing its usability. Other studies could investigate refining the validation process further, by reducing randomness in tasks and/or finding better ways to measure usability from tasks beyond just time and confidence.

Recommendations

Based on findings of this study, there are several recommendations that could be considered and implemented in future research to gather deeper insight and increased consistency.

Firstly, findings suggests that structured tasks provide valuable measurable data on how users interact with a given tool and promote broader use. Future tasks could benefit from an emphasis on making instructions as clear as possible by not assuming any prior knowledge or understanding accompanied by less ambiguous wording. In addition, in person testing through tasks may benefit the overall results as examining participants use of the tool in real time gives insights that users may not relay through surveys, as suggested by Lightbown (2015).

In addition, longer completion times did not reflect the difficulty experience by the user to use the tool as tasks were and could be open ended. Future studies could implement timed tasks, or timed completion in addition to informing participants to complete the task as fast as possible. Furthermore, discrepancies can be caused by intentional randomness within a functionality, whilst testing this can be removed to offer identical experiences for participants.

Additionally, the tools overall usability saw improvement accompanied by a higher complexity reflecting the benefits of user-based development. Continued iterative development using user feedback is likely to contribute to the tools perceived usability.

Lastly, the evaluation of participant feedback indicated a desire for visual feedback on actions taken by users especially when functionality returns errors, this allows users to understand that something is wrong and can potentially guide the user to a solution; if a solution could not be found users are more content knowing an error is occurring then not. Furthermore, when implementing widely used features, participants preferred them to follow standard conventions so that it meets their expectations e.g. Undo is standardised as the CTRL + Z shortcut and not CTRL + B. This should be considered as not only does it improve the tools overall usability it makes development simpler due to using established features rather then redesigning the wheel. This study does however, contribute to the growing body of research evaluating the effectiveness of usability based development for terrain generation and development tools.

References

Avouris, N. (2001) *An Introduction to Software Usability*. Available at: https://hci.ece.upatras.gr/wp-content/uploads/publications/2001(C44)An%20Introduction%20to%20Software%20Usability.pdf.

Bangor, A., Kortum, P.T. and Miller, J.T. (2008) 'An Empirical Evaluation of the System Usability Scale', *International Journal of Human-Computer Interaction*, 24(6), pp. 574–594. Available at: https://doi.org/10.1080/10447310802205776.

Bangor, A., Kortum, P. and Miller, J. (2009) 'Determining what individual SUS scores mean: adding an adjective rating scale', *Journal of Usability Studies archive*, 4(3), pp. 114–123. (Accessed: 24 February 2025)

Blomqvist, S. and Björn Detterfelt (2020) 'Real Time Integrated Tools for Video Game Development :a usability study'. Available at: https://www.diva-portal.org/smash/get/diva2:1447408/FULLTEXT01.pdf

Borg, M. *et al.* (2020) 'Video Game Development in a Rush: A Survey of the Global Game Jam Participants', *IEEE Transactions on Games*, 12(3), pp. 1–1. Available at: https://doi.org/10.1109/tg.2019.2910248

Brooke, J. (1995) *SUS: a Quick and Dirty Usability Scale*, *ResearchGate*. Available at: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale. (Accessed: 24 February 2025)

Brooke, J. (2013) 'SUS: A Retrospective', 8, pp. 29–40. Available at: http://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Brooke_February_2013.pdf. (Accessed: 24 February 2025)

Christopoulou, E. and Xinogalos, S. (2017) 'Overview and Comparative Analysis of Game Engines for Desktop and Mobile Devices', *International Journal of Serious Games*, 4(4). Available at: https://doi.org/10.17083/ijsg.v4i4.194.

Cornut, O. (2019) *ocornut/imgui*, *GitHub*. Available at: https://github.com/ocornut/imgui. (Accessed: 25 February 2025)

Despa, M. (2014) 'Comparative study on software development methodologies', *Database Systems Journal* [Preprint], (3). Available at: https://dbjournal.ro/archive/17/17_4.pdf. (Accessed: 24 February 2025)

Dow, S. *et al.* (2005) 'Wizard of Oz Support throughout an Iterative Design Process', *IEEE Pervasive Computing*, 4(4), pp. 18–26. Available at: https://doi.org/10.1109/mprv.2005.93.

Følstad, A. Users' design feedback in usability evaluation: a literature review. *Hum. Cent. Comput. Inf. Sci.* **7**, 19 (2017). https://doi.org/10.1186/s13673-017-0100-y

Hellstén, T. (2019) 'Iterative design of mobile game UX: Design defined by a target group', *Theseus.fi* [Preprint]. Available at: http://www.theseus.fi/handle/10024/167390. (Accessed: 24 February 2025)

Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A. (2007) 'Toward a definition of mixed methods research', *Journal of Mixed Methods Research*, 1(2), pp. 112–133. Available at: https://journals.sagepub.com/doi/epdf/10.1177/1558689806298224.

Kasurinen, J., Strandén, J.P. and Smolander, K. (2013) 'What do game developers expect from development and design tools?', *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering - EASE '13* [Preprint]. Available at: https://doi.org/10.1145/2460999.2461004.

Kirakowski, J. and Corbett, M. (1993) 'SUMI: the Software Usability Measurement Inventory', *British Journal of Educational Technology*, 24(3), pp. 210–212. Available at: https://doi.org/10.1111/j.1467-8535.1993.tb00076.x.

Lewis, J.R. (1992) 'Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ', *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 2, pp. 1259–1263. Available at: https://www.researchgate.net/publication/230786756_Psychometric_evaluation_of_the_post-study_system_usability_questionnaire_The_PSSUQ (Accessed: 25 October 2024).

Lewis, J.R. (2018) 'Measuring Perceived Usability: The CSUQ, SUS, and UMUX', *International Journal of Human–Computer Interaction*, 34(12), pp. 1148–1156. Available at: https://doi.org/10.1080/10447318.2017.1418805.

Lightbown, D. (2015) *Designing the User Experience of Game Development Tools*. CRC Press.

Macklin, C. and Sharp, J. (2016) *Games, design and play : a detailed approach to iterative game design*. Boston, Ma ; San Francisco, Ca: Addison-Wesley.

Mcdonald, N. (2020) - *Simple Particle-Based Hydraulic Erosion* Available at: https://nickmcd.me/2020/04/10/simple-particle-based-hydraulic-erosion/.

Microsoft (2024) *GitHub - microsoft/DirectXTex: DirectXTex texture processing library*, *GitHub*. Available at: https://github.com/microsoft/DirectXTex. (Accessed: 25 February 2025)

Moran, K. (2018) *Writing Tasks for Quantitative and Qualitative Usability Studies*, *Nielsen Norman Group*. Available at: https://www.nngroup.com/articles/test-tasks-quant-qualitative/

Orfanou, K., Tselios, N. and Katsanos, C. (2015) 'Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale', *The International Review of Research in Open and Distributed Learning*, 16(2). Available at: https://doi.org/10.19173/irrodl.v16i2.1955.

Patel, R. *et al.* (2020) 'Why Software is Not Accessible: Technology Professionals' Perspectives and Challenges', *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* [Preprint]. Available at: https://doi.org/10.1145/3334480.3383103

Peck, J. (2021) *Auburn/FastNoiseLite*, *GitHub*. Available at: https://github.com/Auburn/FastNoiseLite. (Accessed: 25 February 2025)

Putnam, C. *et al.* (2012) 'How do professionals who create computing technologies consider accessibility?', *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12* [Preprint]. Available at: https://doi.org/10.1145/2384916.2384932.

Sauro, J. (2022) *Does Prior Experience Affect Perceptions Of Usability? – MeasuringU*, *Measuringu.com*. Available at: https://measuringu.com/prior-exposure/ (Accessed: 19 February 2025).

Savage, P. (1996) 'User interface evaluation in an iterative design process', *Conference companion on Human factors in computing systems common ground - CHI '96* [Preprint]. Available at: https://doi.org/10.1145/257089.257335.

Schrepp M., Kollmorgen J., Thomaschewski J. (2023) *'A Comparison of SUS, UMUX-LITE, and UEQ-S - JUX JUX - The Journal of User Experience'*. Available at: https://uxpajournal.org/sus-umux-lite-ueq-s/ (Accessed: 25 February 2025)

Syoyo (2025) *GitHub - syoyo/tinygltf: Header only C++11 tiny glTF 2.0 library*, *GitHub*. Available at: https://github.com/syoyo/tinygltf (Accessed: 19 February 2025).

Toftedahl, M. and Engström, H. (2019) *A Taxonomy of Game Engines and the Tools that Drive the Industry*. Available at: https://www.diva-portal.org/smash/get/diva2:1352554/FULLTEXT01.pdf.

Venkatesh, V., Brown, S.A. and Bala, H. (2013) 'Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems', *MIS Quarterly*, 37(1), pp. 21–54. Available at: https://www.jstor.org/stable/43825936. (Accessed: 24 February 2025)

Vohera, C. *et al*. (2021) *Game Engine Architecture and Comparative Study of Different Game Engines*, *IEEE Xplore*. Available at: https://doi.org/10.1109/ICCCNT51525.2021.9579618.

Volioti, C. *et al*. (2024) 'Codeless3D: Design and Usability Evaluation of a Low-Code Tool for 3D Game Generation', *IEEE Transactions on Games*, pp. 1–12. Available at: https://doi.org/10.1109/tg.2024.3424894.

Weber, T., Alois Zoitl and Heinrich Hubmann (2019) 'Usability of Development Tools: A CASE-Study'. Available at: https://doi.org/10.1109/models-c.2019.00037

Appendices

Appendix 1 – Survey 1

1. Have you read the information sheet? *

◯ Yes

◯ No

2. I have read and understood the information sheet. I have been given the opportunity to ask questions, and I have had any questions answered satisfactorily. I understand that my participation in this study is entirely voluntary and that I can withdraw at any time without having to give an explanation.I consent that data collected could be analysed and used to inform the project and understand that all data will be presented anonymously. I agree that data will only be used for this project although the data may also be audited for quality control purposes. All data will be sorted safely on a password protected computer (electronic data), until the project finishes at which point it will be destroyed. I understand that I can withdraw my data from the project up to the point where the data has been analysed without having to give an explanation. *

◯ I hereby give consent to take part in this study.

◯ I do not give consent to take part in this study.

3. What is your occupation? *

Enter your answer

4. What is your age? *

The value must be a number

5. What is your gender? *

Enter your answer

6. Email (Optional if you want to be contacted about the second round of testing and feedback)

Enter your answer

7. How would you describe the computing power of your current device? *

Very low	Low	Moderate	High	Very high
◯	◯	◯	◯	◯

8. Have you used the tool? *

◯ Yes

◯ No

9. System Usability Scale Questionnaire *

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1. I think that I would like to use this system frequently.	○	○	○	○	○
2. I found the system unnecessarily complex.	○	○	○	○	○
3. I thought the system was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this system.	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
8. I found the system very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the system.	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○

10. How would you change the layout to better work for you? *

Enter your answer

11. When using the tool what was your performance like from Very low(Consistently low frame rate, frame drops, stutters and freezes) to Very high(No Freezes or stutters, consistently high frame rate, no performance drops) *

Very low	Low	Moderate	High	Very high
◯	◯	◯	◯	◯

12. If the program stuttered, crashed or had low frame rate, what were you doing when it happened?

Enter your answer

13. How long did it take for you to feel like you understood all the functions of the tool? *

◯ Immediately

◯ After a couple minutes

◯ After an extended period of time

◯ You didnt understand the tool

14. When using the tool was there anything particular that frustrated you or didn't work as you expected and how would you want it to be improved?

Enter your answer

15. Did you experience any bugs while using tool if so what were they?

Enter your answer

16. Is there anything you wish the tool could do that it cant do already? *

Enter your answer

Appendix 2 – Survey 2

1. Have you read the information sheet? *

◯ Yes

◯ No

2. I have read and understood the information sheet. I have been given the opportunity to ask questions, and I have had any questions answered satisfactorily. I understand that my participation in this study is entirely voluntary and that I can withdraw at any time without having to give an explanation.I consent that data collected could be analysed and used to inform the project and understand that all data will be presented anonymously. I agree that data will only be used for this project although the data may also be audited for quality control purposes. All data will be sorted safely on a password protected computer (electronic data), until the project finishes at which point it will be destroyed. I understand that I can withdraw my data from the project up to the point where the data has been analysed without having to give an explanation. *

◯ I hereby give consent to take part in this study.

◯ I do not give consent to take part in this study.

3. What is your occupation? *

Enter your answer

4. What is your age? *

The value must be a number

5. What is your gender? *

Enter your answer

6. Email (Optional if you want to be contacted about the second round of testing and feedback)

Enter your answer

7. Did you participate in the last survey? *

- ◯ Yes
- ◯ No

8. How would you describe the computing power of your current device? *

	Very low	Low	Moderate	High	Very high
	◯	◯	◯	◯	◯

9. Have you used the tool? *

- ◯ Yes
- ◯ No

10. After reading the instructions for Task 1, how confident were you in understanding what you needed to do? *

- ◯ I was completely confident in what I had to do.
- ◯ I was fairly confident in what I had to do.
- ◯ I was neither confident nor unsure of what I had to do.
- ◯ I was somewhat unsure of what I had to do.
- ◯ I was not confident at all in what I had to do.

11. For Task 1, how long did it take you to complete it? *

- ◯ Less than 5 minutes.
- ◯ 5 to 10 minutes.
- ◯ More than 10 minutes.
- ◯ I did not complete the task.

12. After reading the instructions for Task 2, how confident were you in understanding what you needed to do? *

- ◯ I was completely confident in what I had to do.
- ◯ I was fairly confident in what I had to do.
- ◯ I was neither confident nor unsure of what I had to do.
- ◯ I was somewhat unsure of what I had to do.
- ◯ I was not confident at all in what I had to do.

13. For Task 2, how long did it take you to complete it? *

- ◯ Less than 5 minutes.
- ◯ 5 to 10 minutes.
- ◯ More than 10 minutes.
- ◯ I did not complete the task.

14. System Usability Scale Questionnaire *

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly c
1. I think that I would like to use this system frequently.	◯	◯	◯	◯	◯
2. I found the system unnecessarily complex.	◯	◯	◯	◯	◯
3. I thought the system was easy to use.	◯	◯	◯	◯	◯
4. I think that I would need the support of a technical person to be able to use this system.	◯	◯	◯	◯	◯
5. I found the various functions in this system were well integrated.	◯	◯	◯	◯	◯
6. I thought there was too much inconsistency in this system.	◯	◯	◯	◯	◯
7. I would imagine that most people would learn to use this system very quickly.	◯	◯	◯	◯	◯
8. I found the system very cumbersome to use.	◯	◯	◯	◯	◯
9. I felt very confident using the system.	◯	◯	◯	◯	◯
10. I needed to learn a lot of things before I could get going with this system.	◯	◯	◯	◯	◯

15. How would you compare the new version of the tool to the previous version in terms of usability and functionality?

Much worse	Slightly worse	About the same	Slightly better	Much better
○	○	○	○	○

16. How would you change the layout to better work for you? *

Enter your answer

17. When using the tool what was your performance like from Very low(Consistently low frame rate, frame drops, stutters and freezes) to Very high(No Freezes or stutters, consistently high frame rate, no performance drops) *

Very low	Low	Moderate	High	Very high
○	○	○	○	○

18. If the program stuttered, crashed or had low frame rate, what were you doing when it happened?

Enter your answer

19. How does the performance of the new version compare to the previous version?

Much worse	Slightly worse	About the same	Slightly better	Much better
○	○	○	○	○

20. How long did it take for you to feel like you understood all the functions of the tool? *

- ◯ Immediately
- ◯ After a couple minutes
- ◯ After an extended period of time
- ◯ You didnt understand the tool

21. When using the tool was there anything particular that frustrated you or didn't work as you expected and how would you want it to be improved?

Enter your answer

22. Did you experience any bugs while using tool if so what were they?

Enter your answer

23. What brushes / tools would you like to be added to enable you to directly manipulate the terrain? *

e.g. The already added flatten brush

Enter your answer

24. Is there anything you wish the tool could do that it cant do already or ideas for planned additions? *

Planned features are: Undo, Image Saving, Chosing save location and Texturing / Colouring the terrain.

Enter your answer

Appendix 3 – Survey 3

1. Have you read the information sheet? *

   ◯ Yes

   ◯ No

2. I have read and understood the information sheet. I have been given the opportunity to ask questions, and I have had any questions answered satisfactorily. I understand that my participation in this study is entirely voluntary and that I can withdraw at any time without having to give an explanation.I consent that data collected could be analysed and used to inform the project and understand that all data will be presented anonymously. I agree that data will only be used for this project although the data may also be audited for quality control purposes. All data will be sorted safely on a password protected computer (electronic data), until the project finishes at which point it will be destroyed. I understand that I can withdraw my data from the project up to the point where the data has been analysed without having to give an explanation. *

   ◯ I hereby give consent to take part in this study.

   ◯ I do not give consent to take part in this study.

3. What is your occupation? *

   Enter your answer

4. What is your age? *

   The value must be a number

5. What is your gender? *

   Enter your answer

6. Email (Optional if you want to be contacted about the second round of testing and feedback)

   Enter your answer

7. Did you participate in one of the previous surveys? *

   ◯ Yes

   ◯ No

8. How would you describe the computing power of your current device? *

Very low	Low	Moderate	High	Very high
◯	◯	◯	◯	◯

9. Have you used the tool? *

   ◯ Yes

   ◯ No

10. After reading the instructions for Task 1, how confident were you in understanding what you needed to do? *

- I was completely confident in what I had to do.
- I was fairly confident in what I had to do.
- I was neither confident nor unsure of what I had to do.
- I was somewhat unsure of what I had to do.
- I was not confident at all in what I had to do.

11. For Task 1, how long did it take you to complete it? *

- Less than 5 minutes.
- 5 to 10 minutes.
- More than 10 minutes.
- I did not complete the task.

12. After reading the instructions for Task 2, how confident were you in understanding what you needed to do? *

- I was completely confident in what I had to do.
- I was fairly confident in what I had to do.
- I was neither confident nor unsure of what I had to do.
- I was somewhat unsure of what I had to do.
- I was not confident at all in what I had to do.

13. For Task 2, how long did it take you to complete it? *

- Less than 5 minutes.
- 5 to 10 minutes.
- More than 10 minutes.
- I did not complete the task.

14. System Usability Scale Questionnaire *

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly (
1. I think that I would like to use this system frequently.	○	○	○	○	○
2. I found the system unnecessarily complex.	○	○	○	○	○
3. I thought the system was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this system.	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
8. I found the system very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the system.	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○

15. How would you compare the new version of the tool to the previous version in terms of usability and functionality?

Much worse	Slightly worse	About the same	Slightly better	Much better
◯	◯	◯	◯	◯

16. How would you change the layout to better work for you? *

Enter your answer

17. When using the tool what was your performance like from Very low(Consistently low frame rate, frame drops, stutters and freezes) to Very high(No Freezes or stutters, consistently high frame rate, no performance drops) *

Very low	Low	Moderate	High	Very high
◯	◯	◯	◯	◯

18. If the program stuttered, crashed or had low frame rate, what were you doing when it happened?

Enter your answer

19. How does the performance of the new version compare to the previous version?

Much worse	Slightly worse	About the same	Slightly better	Much better
◯	◯	◯	◯	◯

20. How long did it take for you to feel like you understood all the functions of the tool? *

- ◯ Immediately
- ◯ After a couple minutes
- ◯ After an extended period of time
- ◯ You didnt understand the tool

21. When using the tool was there anything particular that frustrated you or didn't work as you expected and how would you want it to be improved?

Enter your answer

22. Did you experience any bugs while using tool if so what were they?

Enter your answer

23. Is there anything you wish the tool could do that it cant do already or ideas for planned additions? *

Planned features are: Undo & Redo, Colouring the terrain based on height.

Enter your answer

Appendix 4 – Survey 4

1. Have you read the information sheet? *

◯ Yes

◯ No

2. I have read and understood the information sheet. I have been given the opportunity to ask questions, and I have had any questions answered satisfactorily. I understand that my participation in this study is entirely voluntary and that I can withdraw at any time without having to give an explanation.I consent that data collected could be analysed and used to inform the project and understand that all data will be presented anonymously. I agree that data will only be used for this project although the data may also be audited for quality control purposes. All data will be sorted safely on a password protected computer (electronic data), until the project finishes at which point it will be destroyed. I understand that I can withdraw my data from the project up to the point where the data has been analysed without having to give an explanation. *

◯ I hereby give consent to take part in this study.

◯ I do not give consent to take part in this study.

3. What is your occupation? *

Enter your answer

4. What is your age? *

The value must be a number

5. What is your gender? *

Enter your answer

6. Did you participate in one of the previous surveys? *

◯ Yes

◯ No

7. How would you describe the computing power of your current device? *

Very low	Low	Moderate	High	Very high
◯	◯	◯	◯	◯

8. Have you used the tool? *

◯ Yes

◯ No

9. After reading the instructions for Task 1, how confident were you in understanding what you needed to do? *

- I was completely confident in what I had to do.
- I was fairly confident in what I had to do.
- I was neither confident nor unsure of what I had to do.
- I was somewhat unsure of what I had to do.
- I was not confident at all in what I had to do.

10. For Task 1, how long did it take you to complete it? *

- Less than 5 minutes.
- 5 to 10 minutes.
- More than 10 minutes.
- I did not complete the task.

11. After reading the instructions for Task 2, how confident were you in understanding what you needed to do? *

- I was completely confident in what I had to do.
- I was fairly confident in what I had to do.
- I was neither confident nor unsure of what I had to do.
- I was somewhat unsure of what I had to do.
- I was not confident at all in what I had to do.

12. For Task 2, how long did it take you to complete it? *

- Less than 5 minutes.
- 5 to 10 minutes.
- More than 10 minutes.
- I did not complete the task.

13. After reading the instructions for Task 3, how confident were you in understanding what you needed to do? *

- I was completely confident in what I had to do.
- I was fairly confident in what I had to do.
- I was neither confident nor unsure of what I had to do.
- I was somewhat unsure of what I had to do.
- I was not confident at all in what I had to do.

14. For Task 3, how long did it take you to complete it? *

- Less than 5 minutes.
- 5 to 10 minutes.
- More than 10 minutes.
- I did not complete the task.

15. If you have viewed the tutorial did you find it useful.

Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
○	○	○	○	○

16. Are there improvements you would like to see to the tutorial, if so what would they be?

Enter your answer

17. System Usability Scale Questionnaire *

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly c
1. I think that I would like to use this system frequently.	○	○	○	○	○
2. I found the system unnecessarily complex.	○	○	○	○	○
3. I thought the system was easy to use.	○	○	○	○	○
4. I think that I would need the support of a technical person to be able to use this system.	○	○	○	○	○
5. I found the various functions in this system were well integrated.	○	○	○	○	○
6. I thought there was too much inconsistency in this system.	○	○	○	○	○
7. I would imagine that most people would learn to use this system very quickly.	○	○	○	○	○
8. I found the system very cumbersome to use.	○	○	○	○	○
9. I felt very confident using the system.	○	○	○	○	○
10. I needed to learn a lot of things before I could get going with this system.	○	○	○	○	○

18. If you participated previously how would you compare the new version of the tool to the previous version in terms of usability and functionality?

Much worse	Slightly worse	About the same	Slightly better	Much better
○	○	○	○	○

19. How would you change the layout to better work for you? *

Enter your answer

20. When using the tool what was your performance like from Very low(Consistently low frame rate, frame drops, stutters and freezes) to Very high(No Freezes or stutters, consistently high frame rate, no performance drops) *

Very low	Low	Moderate	High	Very high
○	○	○	○	○

21. If the program stuttered, crashed or had low frame rate, what were you doing when it happened?

Enter your answer

22. How does the performance of the new version compare to the previous version?

	Much worse	Slightly worse	About the same	Slightly better	Much better
	◯	◯	◯	◯	◯

23. How long did it take for you to feel like you understood all the functions of the tool? *

◯ Immediately

◯ After a couple minutes

◯ After an extended period of time

◯ You didnt understand the tool

24. When using the tool was there anything particular that frustrated you or didn't work as you expected and how would you want it to be improved?

Enter your answer

25. Did you experience any bugs while using tool if so what were they?

Enter your answer

26. Is there anything you wish the tool could do that it cant do already or ideas for the planned additions? *

Planned features are: Colouring the terrain based on height, Dynamic fill tool.

Enter your answer