A dashboard and Learning Analytics to provide immediate feedback for game-based learning

Vanshika Agarwal

April 30, 2020

A Final Year Project submitted in partial fulfilment of the requirements for the degree of MAI (Computer Engineering)

Supervised by: Dr. Vincent Wade
Declaration

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism ‘Ready Steady Write’, located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

Signed: Vanshika Agarwal

Date: 30/04/2020
Abstract

Training deficits in ‘soft skills’ - interpersonal or non-technical skills - have long been lamented by tutors and employers despite longstanding evidence of their importance. Virtual environments, combined with online learning management and reporting platforms, offer potential for addressing this gap through affordable and scalable simulations. Even though serious games have now been increasingly adopted, there are very few software tools that allow efficient applications for their assessments. This research aims to create a reporting platform for one such serious game called Gaming for Peace (GAP) and to explore practical issues encountered by development teams working in this area.
Acknowledgements

I would like to express my heartfelt gratitude and sincere thanks to Dr. Vincent Wade, Chair of Computer Science. He not only supervised me but helped me grow with this research. He genuinely cared about the work and were very responsive to all my queries. His constant guidance and motivation in the unprecedented and stressful times of Coronavirus made the research to what it is today.

I am grateful to Matthew Nicholson, research engineer at ADAPT Centre for his continued time and support from getting me on-boarded till the completion of my research.

I would like to sincerely thank my parents and siblings for their never ending love and support throughout this journey. I’d like to especially thank my father for taking out time from his busy routine to help me conduct a user trial that was an integral part of my thesis but had to be abandoned due to a Coronavirus. He was the only user I was able to test my research on.

Lastly, I’m eternally grateful to my few amazing friends who stuck by my side throughout the university days and made this time worth remembering.
## Contents

1 Introduction ................................. 1
   1.1 Background .............................. 1
   1.2 Motivation ................................ 2
   1.3 Requirements ............................ 3
   1.4 Structure ................................ 3

2 State of the Art ............................. 5
   2.1 Gaming for peace (GAP) ................. 5
   2.2 Kirkpatrick’s Model ...................... 7
   2.3 Learning assessment in games .......... 9

3 Background and Design ..................... 10
   3.1 Game Overview ........................... 10
      3.1.1 Game mechanics ...................... 11
   3.2 Design Requirements ..................... 15
   3.3 Design Challenges ....................... 16
   3.4 Technologies used ....................... 18
      3.4.1 Data Preparation ..................... 19
      3.4.2 Data Visualisation ................... 19
      3.4.3 Deployment ........................... 19

4 Method and Implementation ............... 21
   4.1 Software development lifecycle ........ 21
      4.1.1 Data collection and preparation ..... 22
      4.1.2 Data modelling ....................... 22
      4.1.3 Evaluation ............................ 28
      4.1.4 Deployment ............................ 30
      4.1.5 Monitor ................................ 30
   4.2 Workflow .................................. 30
      4.2.1 Workflow between the game and the assessment engine ... 31
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Kirkpatrick Model: Four Levels of Learning Evaluation</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>A screen-grab from the Gaming for Peace simulation training software</td>
<td>11</td>
</tr>
<tr>
<td>3.2</td>
<td>A screen-grab of the <strong>monologue</strong> from the GAP simulation training software</td>
<td>12</td>
</tr>
<tr>
<td>3.3</td>
<td>A screen-grab of a <strong>conversation</strong> between characters from the GAP simulation training software</td>
<td>12</td>
</tr>
<tr>
<td>3.4</td>
<td>A screen-grab of an <strong>exploration</strong> scene from the GAP simulation training software</td>
<td>13</td>
</tr>
<tr>
<td>3.5</td>
<td>A screen-grab of <strong>feedback</strong> generated after each checkpoint in GAP simulation training software</td>
<td>14</td>
</tr>
<tr>
<td>4.1</td>
<td>Software development life cycle (SDLC) of the game assessment system</td>
<td>21</td>
</tr>
<tr>
<td>4.2</td>
<td>A JSON file for a player’s gameplay</td>
<td>22</td>
</tr>
<tr>
<td>4.3</td>
<td>A player’s ranking in communication competency throughout the scenes in the role of a female military liaison officer</td>
<td>23</td>
</tr>
<tr>
<td>4.4</td>
<td>Inter-Cohort performances</td>
<td>25</td>
</tr>
<tr>
<td>4.5</td>
<td>Arithmetic scores of a game play</td>
<td>26</td>
</tr>
<tr>
<td>4.6</td>
<td>Leaderboard</td>
<td>27</td>
</tr>
<tr>
<td>4.7</td>
<td>Performance comparison using Kernel density estimate on absolute scores for returning players</td>
<td>28</td>
</tr>
<tr>
<td>4.8</td>
<td>Chapter performance per competency</td>
<td>29</td>
</tr>
<tr>
<td>4.9</td>
<td>A waterfall chart to visualise learning rate</td>
<td>29</td>
</tr>
<tr>
<td>4.10</td>
<td>A performance comparison between plays for a single player</td>
<td>30</td>
</tr>
<tr>
<td>4.11</td>
<td>Lifecycle of the game assessment system</td>
<td>31</td>
</tr>
<tr>
<td>4.12</td>
<td>Workflow between the GAP game and the assessment system</td>
<td>32</td>
</tr>
<tr>
<td>4.13</td>
<td>Navigation bar on the left with the player’s arithmetic summary beside it</td>
<td>33</td>
</tr>
<tr>
<td>4.14</td>
<td>Choices chosen in the gameplay with their corresponding scores are displayed</td>
<td>34</td>
</tr>
</tbody>
</table>
4.15 A screen-grab of the dashboard describing all the sub learning objectives associated with a scene and the best fit solution in it . . . . . . . . 35
4.16 A feedback section in the assessment dashboard for end-user opinions . 35
List of Tables

3.1 The approach to adding weights based on the users’ data . . . . . . . . . . . 18
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORDIS</td>
<td>Community Research and Development Information Service</td>
</tr>
<tr>
<td>CPPB</td>
<td>Conflict Prevention and Peace Building</td>
</tr>
<tr>
<td>CSDP</td>
<td>Common Security and Defence Policy</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Values</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>GAP</td>
<td>Gaming for Peace</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical user interface</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hyper Text Transfer protocol</td>
</tr>
<tr>
<td>ICO</td>
<td>Interactive Content Object</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-Governmental Organisation</td>
</tr>
<tr>
<td>SDLC</td>
<td>Software development life cycle</td>
</tr>
<tr>
<td>UNSCR</td>
<td>United Nations Security Council Resolutions</td>
</tr>
</tbody>
</table>
1 Introduction

Research tells us that soft skills\(^1\), such as leadership skills, negotiation skills, communication skills and cultural awareness skills both predict and produce future career and life success [5], [6]. However, the development of these skills are often overlooked, both during the training of future professionals and in the workplace itself, where achievement tests tend to miss the “personality traits, goals, motivations, and preferences” [5] that denote the acquisition of soft skills [7]. Educating individuals on building effective interpersonal skills is critically important for any industry and for individuals’ careers. This is especially challenging as it relates to teamwork and soft skills such as organisation, project management, teamwork ethics and practices, collaborative and communication skills and practices, as well as the ability to effectively use the required tools [8]. Soft skills are interpersonal skills that encompass both personality traits and abilities that can be learnt. They, by definition, cannot be taught but can certainly be learnt and developed (not by demonstration but by participation, discussion and practice) [9].

1.1 Background

A game called GAP, as part of an EU project Gaming for Peace\(^2\) [10] was developed to train military, police and civilian peacekeepers deployed in conflict zones to practise and learn soft skills. The project proposed that Serious Games\(^3\) offer a 21st century environment within which Conflict Prevention and Peace Building (CPPB) personnel could experience scenarios through role-playing in their own organisation and by role-playing people from other organisations. In doing so, the military peacekeeping personnel increase their understanding, creativity and ability to communicate and col-

\(^1\)Soft skills are a combination of people skills, social skills, communication skills, character or personality traits, attitudes, career attributes [2], social intelligence and emotional intelligence quotients, among others, that enable people to navigate their environment, work well with others, perform well, and achieve their goals with complementing hard skills [3]. Source: [4].

\(^2\)CORDIS | European Commission, 2020. Available at: https://cordis.europa.eu/project/id/700670

\(^3\)Serious Games are games designed with a primary purpose other than pure recreation
laborate with other organisations in the network, which is a CPPB mission. In particular, 3 soft skill competencies are addressed in the game, namely, communication skill, cultural awareness skill and gender awareness skill. It is an immersive role-playing game developed for peacekeepers. The game has 2 roleplays; one as a female community liaison officer in the military where the user goes on patrols and communicates and coordinates activities with locals of the town and solve problems. This roleplay is fictionally set for a period of 6 days spread over on 17 days story-line. The game then moves on to the second role, that is of a male police liaison officer who is a mentor to the local policemen giving advice and lessons on peacekeeping strategies. This roleplay is set for 6 days on a 14 days story-line. Both roles have a few overlapping days to witness the working of other peacekeepers in the same situation. The game, at the end, provides the user with a 5-star rating on their performance in all three interpersonal skills. However, no actionable information is presented to improve or understand the score. Therefore, a number of learning challenges are posed by the game that could be addressed to improve players’ performances and develop their soft skills.

1.2 Motivation

One of the problems with current military training is the high risk of skill diminishment which arises from insufficient opportunities to apply the learning in real peacebuilding situations after the training, as emergency incidents involving peacekeeping missions are relatively uncommon and can be expensive to recreate in real training exercises [11]. Suppose a user played the game and received a rating of 3/5 on one of the 3 competencies - this standalone value is insufficient to the user as he will not know if it is good, average or poor with respect to his fellow colleagues’ performances. If a user scores well in a scene, they have no concrete measure to know if everyone scored that well because it was an "easy" scene or the user indeed demonstrated skills relevant for peacebuilding. On the other hand, if the game developers provide a user with scores and answers in every scene, it will be information overload and the user will be unable to pick up on any valuable information.

Lastly, the game has over 550 decision points which lead to the generation of the final star rating. Each scene can have multiple decision points and is impossible for the player to understand them all during the game. The requirement, hence, is to let a user play the game as is but later have a platform that reflects on all those decision points. How can the user be informed about all these learning objectives? A middle ground is required to enable users to explore their performance in every scene and learn about the alternative answers.
1.3 Requirements

This thesis aims to assess the soft skills portrayed by a trainee/user. A 5-star rating is in place which recognises all questions as equal and therefore, takes the mean of scores in each non-zero\(^4\) question. A more thorough mechanism is required which considers the difficulty level of each question and calculates the final score accordingly because "not all questions are created equal". The results also need to include graphical representations of the trends and advancements in the various competencies a user showcased throughout its gameplay because of the sheer amount of data (scenes) a user needs to analyze.

This analysis will also be extended to send aggregated results to tutors and trainees - with the idea to assess individual and divisional performances and to compare with and learn from the performance of other users. This will facilitate the tutors to lead the development of skills which are deemed to fall short of expectations for the given scenario. The three competencies - Communications, Culture and Gender - are further divided into subcategories which may be touched upon as an enhancement feature to the competency graphs.

Further, the research consists of building a suggestion box that targets the scoring structure to provide an option to appeal. The player can argue why the chosen option was the best choice (or the appropriate choice) given the circumstances. This data can be further utilised to improve the scoring mechanism either manually or using advanced machine learning techniques like Natural Language Processing (NLP).

1.4 Structure

Following is the structure of the report-

**Chapter 2** discusses the current state of the art describing the commonly used methods for learning evaluation and a description of the game on which the research is based upon.

**Chapter 3** presents the game design and the design requirements and challenges faced in building the feedback system.

**Chapter 4** describes the methodology and implementation of the design proposed.

**Chapter 5** evaluates end user-trial and if all requirements proposed in the design were fulfilled.

\(^4\)A question is regarded as non-zero (or relevant) with a minimum score of 1 in any of the three competencies.
Chapter 6 concludes the findings and suggests future work that could be done following the findings of the work done in this project.
2 State of the Art

This chapter introduces the state of the art in game based learning and, in particular, provides an initial overview of the Gaming For Peace immersive game and identifies key assessment and pedagogical models being used as a basis to evaluate learners. These theories will form requirements for the design of the learner evaluation within the immersive game, GAP.

2.1 Gaming for peace (GAP)

GAP is an EU H2020 Framework Programme for Research and Innovation project and the main goal of GAP was to develop a curriculum in relevant soft skills (including cooperation, communication, negotiation, gender and cultural awareness) for personnel from diverse organisations working in the fields of conflict prevention, peace-building and peacekeeping operations. This curriculum was embedded into an immersive online role-playing game designed to train peace-keepers in soft skills like communication, gender awareness and cultural awareness which are crucial to achieve cooperation in complex peace keeping missions. Peacekeeping missions comprise of diverse organisations (like military, police forces and non-government organisations), nations, cultures (nationality, ethnicity and religion) and a diversity of genders.

The work performed by the GAP Consortium partners identified the need for soft skills by researching the relevant training and literature, and through conducting interviews with peacekeepers. This consortium enabled the developers to produce a variety of challenging and authentic scenarios based upon events which have both contemporary and cultural relevance, and which incite the trainee/user to practice and improve their soft skills in an online ‘safe space’\(^1\). The intention was for these soft skills to be tested and refined in both abstract and defined circumstances to showcase both the flexibility of the skills already possessed by the individual as well as to provoke the development of skills which are deemed to fall short of expectations for

\(^1\)Safe space is an environment in which a person or category of people can feel confident that they will not be exposed to discrimination, criticism, harassment, or any other emotional or physical harm
the given scenario. The focus is not simply on the reinforcement of existing skills but on the recognition and highlighting of areas in which one may be able to up-skill via employing new behaviour and strategies available through the digital setting of GAP. These skills are then intended to be employed in reality with the performance in the online scenarios being bench-marked against existing, recognised international standards. It was intended that such an online setting would also allow for the individual to experiment with their soft skills to potentially encourage more robust and creative applications for such skills when interacting with police, peacekeepers, and the civilian populaces of an array of given cultural, historical, or gendered backgrounds. Gender and cultural awareness will be driven by the experience of role-playing someone of a different gender or cultural background, but the curricula can also include relevant readings and resources, links to relevant UNSCR\textsuperscript{2} resolutions, theories of difference, feminist theories and gender-aware empirical evidence. The instruments for the testing of the soft skills before and after serious game play are bench-marked against international standards.

Research indicated [12] there was a need to ensure that the end-users are afforded a clear and measurable metric of skill standards provided via the creation of a Skills Passport which entails the transfer and recognition of skills acquired through the GAP program. The purpose of this deliverable was to provide transparent and defined outcome criteria and skills competency analysis which has been tracked and compared against International Skills Standards with European legislation. These strategies are embedded within the GAP game, providing stakeholders with the ability to assess progress for each competency. The game is built with particular reference to the individual’s capabilities in decision-making under high stress environments and interaction in a culturally and organisationally diverse mission.

The GAP consortium prioritised some learning objectives across communication, culture awareness and gender awareness, and designed the GAP curriculum to focus on those objectives. These objectives are summarised in Appendix A.1. The GAP consortium developed two scenarios for the game which allowed multiple perspectives to be taken across these learning outcomes. Scenario one took the perspective of a military person while scenario two took the perspective of a police person. The two scenarios used to investigate and evaluate the three areas identified in the learning outcomes are developed with an additional account on how the history, politics and power of individuals and organisations might have shaped previous peacekeeping missions and, hence, providing realism in the game.

An important aspect of the GAP project is to ensure that the performance of playthroughs

\textsuperscript{2}United Nations Security Council Resolutions (UNSCR)
can be measured against well-defined Learning Objectives for communication, culture awareness and gender awareness. This dissertation will outline a general criteria that makes a learning outcome more or less relevant for a player/tutor, keeping the game’s priority on how well the learning outcome incorporates aspects related to trust, empathy, cooperation, communication, cultural awareness and gender awareness. Generally, learning outcomes related to having experiences - as a character, in a place, etc - are better retained, while those related to quizzing a player’s knowledge are less so.

The objective of the research described in this report is a dashboard and Learning Analytics to provide immediate feedback for game-based learning, is to improve the learning in the game via improving assessment reporting for e.g. educators worldwide are now increasingly using learning analytics as a key aspect in their design to inform the student learner about their performance in a more meaningful way but to also to generate insights on how better those learners can learn. Learning analytics is coming to the aid of learners to try and be much more accurate in terms of their performance and how they might improve [13].

2.2 Kirkpatrick’s Model

The most widely accepted model for evaluating training is the Kirkpatrick model [14].

This model was developed by Dr. Donald Kirkpatrick (1924 – 2014) in the 1950s. The model as shown in Figure 2.1 can be implemented before, throughout, and following training to show the value of training to the organisation [15].

The Kirkpatrick model divides learning into different layers and their levels.

Level one, which is Reaction, evaluates how individuals react to the training model by asking questions that established the trainees’ thoughts.

Level two is Learning. Evaluating at this level is meant to gauge the level participants have developed in expertise, knowledge, or mindset. Exploration at this level is far more challenging and time-consuming compared to level one.

Level three is Transfer. This level analyses the differences in the participant’s behaviour at work after completing the program. Assessing the change makes it possible to figure out if the knowledge, mindset, or skills the program taught are being used the workplace.

Level four or the top-level is the Result stage. It is the in-situ change in the organisation. Commonly regarded as the primary goal of any program, level
four determines the overall success of the training model by measuring factors such as lowered inter/intra organisational conflicts, reduced resistance rates, etc.

As outlined by this system, evaluation needs to start with level one, after which as time and resources will allow, should proceed in order through levels two, three, and four. Data from all of the previous levels can be used as a foundation for the following levels’ analysis. As a result, each subsequent level provides an even more accurate measurement of the usefulness of the training course, yet simultaneously calls for a significantly more time-consuming and demanding evaluation [14], [15].

Undoubtedly, the most widely used and in-demand method for the assessment of training in businesses nowadays is Kirkpatrick’s system based on the four levels as guidelines. The Kirkpatrick model has been used for over 30 years by many different types of organisations as the major system for training evaluations. Kirkpatrick could also be applied to peacekeeping and peacebuilding training evaluation. A general idea of the 4 stages peacekeeping organisations could follow is presented. Level 1 being evaluation of reaction of the military, police and NGOs to their suggested training courses and structures. Level 2 could evaluate peacekeepers’ training in classrooms as well as in practical assessments and Level 3 could focus on evaluation of performance of soft skills ‘in theatre’ (i.e. on the CPPB missions). Lastly, Level 4 could focus on evaluating an overall organisational performance and an in-situ change through peacekeeping missions and the role played by the (trained) peacekeepers in shaping
the culture of their organisation. It is evident that Kirkpatrick’s model can be applied to every training setting and can make a positive impact to the overall practice of training evaluation [15].

The aim of this study was to measure and assess behavioural change from training players in a simulated environment i.e. the GAP game. It is thus the third level of the Kirkpatrick model. The reaction and learning levels of the Kirkpatrick model (levels 1 and 2) have already been fulfilled, where level one was evaluated by the peacekeeping organisations during training and level two was concluded in the GAP game development team at the GAP Soft Skills Summit, held in January 2019. Building a tool that transitions from level 2 to level 3 would involve gathering information on the participants’ reaction to using GAP game and the tabletop exercise with the tutors. This data will then be used to develop a comprehensive reporting mechanism that provides subtle insights on the player’s performances. It would be impracticable to evaluate at Kirkpatrick’s level 4, which focuses on job behaviour and benefits to the society. It is very difficult to measure learning at this stage as it would require the participants to attend real peacekeeping missions and therefore, has scope to be implemented in the future.

2.3 Learning assessment in games

As part of the research, a number of games that induced game-based learning were looked at [16], [17], [18], [19] etc. including research models presented for the learning of soft skills [20]. Soon, a pattern in those papers and proposals was realised; most of the research focused upon the games and the immersive role-playing scenarios, whereas focus on a feedback mechanism/dashboard was almost non-existent. Even for the proposals for which the focal point was feedback [21], they often seemed to lack a multi-approach, interactivity and an all-inclusive feedback mechanism.

A great percentage of them focus on game design and on the statistical analysis [22]. Very few amongst them have contained a dashboard from which the learner could actually learn from and help change the behaviour. Therefore, a novel approach in this research is to create an assessment feedback as a learning tool within the game.
3 Background and Design

This chapter gives an overview of the data and technologies that have been used in this project. Design elements like the working of the GAP game, requirements for building a feedback assessment system, its challenges and a background of the main technologies used to build the assessment system are discussed here.

3.1 Game Overview

Before describing the design aspect of this project, a description of how the game works is presented. GAP is a role-playing game where the player is immersed in a curriculum with close reference to the experiences of those who have participated in overseas deployment for peacekeeping missions. The game pays close attention to the realism of the scenarios both on and off the military base and how the interactions with international colleagues as well as locals should be conducted. The game provides the player with an avatar and its background at the initial stages of the game. The player is supposed to completely submerge into the avatar and proceed with the events being displayed on the screen. The scenes are interactions between people, prompting the user to converse by choosing the best-fit answer (from a multiple-choice rubric) according to their thought process. According to Figure 3.1 the avatar is given a narrative on the bottom of the screen and asked to respond from one of the three available answers.
3.1.1 Game mechanics

The game’s functionality is divided into three main gameplay modes. Each mode gives the player access to a subset of the carefully designed game curriculum. Each mode has a different purpose in the game and has its own user interface (UI) that facilitates the interaction with the user. The game is in exactly one play mode at any given point in the game and may changeover to different modes either due to player interaction in the game or due to the story progression in the game. Below are the main gameplay modes defined in the GAP curriculum:

A. Inner Monologue

The inner monologue mode as in Figure 3.2 is a combination of a narrative and a soliloquy that a player is exposed to when in the game. The game starts with this particular play mode, informing the player about his/her role in the “safe” peacekeeping mission i.e. the game. It is also used to convey the thoughts and feelings of this avatar, bringing it to life in the player’s mind. This mode is entirely textual, so the text is entered through a slow animation to encourage full text reading rather than skimming. The lack of visuals in this mode is deliberate to keep the player in a receptive, imaginative cognitive space, rather than an active, and exploring one.

B. Conversation

This play mode allows the player to converse with the other characters in the game. They could be characters from the same organisation as of the player’s role, or from a different organisation or locals of the area. Figure 3.3 is an example of how a con-
As we bounced back through the streets of Najara on our return to base, I thought through everything that had happened in the market today.

I was beginning to see how complicated life out here could be.

But even so, Najara seemed more normal than I had imagined.

Women in the market, kids playing in the streets, men smoking and drinking coffee.

Next

Figure 3.2: A screen-grab of the monologue from the GAP simulation training software

conversation proceeds with dialogue/s from an internal character who is highlighted to distinguish from other characters present onscreen, and waits for a response from the player through selection of one of the multiple choices displayed at the centre of the screen.

Figure 3.3: A screen-grab of a conversation between characters from the GAP simulation training software

*Mattie leans back and waits for your assessment.*
C. Exploration

This mode is a pan over. It is a 180° view of an area (e.g. market scene) with multiple characters flashing to converse with. The player gets the opportunity to analyse the visuals, assess the mentors advice and wander around to talk to as many people as they may please. Exploration gives an opportunity to make decisions based on context, situation, culture and gender.

![Figure 3.4: A screen-grab of an exploration scene from the GAP simulation training software](image)

D. Feedback

Since the game is fairly long, game developers have divided the game into sections. At the end of each section a feedback report is generated for review. If the player may wish they can open this feedback, as shown in Figure 3.5 and review their selected answers. This report also provides a note on the correctness of the chosen answer through very good, good and poor adjectives attached to each of the three soft skill assessed. Then it also provides an overall star rating on the three competencies and a metric of the percentage of other players who selected the same or different choice. It is evidently an interesting way to learn the like mindedness of oneself amongst their fellow peacekeepers. However, nothing apart from getting a general idea of one’s performance, can be inferred from this assessment report.
Figure 3.5: A screen-grab of feedback generated after each checkpoint in GAP simulation training software
3.2 Design Requirements

While military personnel/p police are considered to be responsible for building their knowledge on soft and hard skills, their learning is also understood to depend on training institutions and staff providing learning and environmental conditions that stimulate, encourage and value trainer involvement and participation. Since their primary job involves finding the difficult path from conflict to peace in war/conflict zones, providing environmental conditions like a real battlefield for learning purposes would be impractical.

Thus, the GAP game was developed. However, the primary objective of learning still remains unfulfilled. How would the trainee (player) know how he/she performed? Did he/she gain any value at the end of the game or not? This induces the first requirement -

1. Development of a reporting mechanism for the trainee to reflect on his/her performance that spurs game-based learning.

Previous studies indicate that immediate feedback has the potential to improve learning and performance [23], [24]. Epstein et al. demonstrated in their work on academia that students given immediate feedback on the correctness of their answers performed better than their colleagues who were given feedback on evaluations after 1 day or 1 week of the initial assessment. This research argues that immediacy for better learning is valid for GAP training as well. In order to achieve the immediacy of the assessment feedback, the second requirement arises -

2. To develop a live system/dashboard that provides immediate feedback to the learners after a gameplay.

Further, to enable cognitive learning that is active, constructive, and long-lasting; the assessment system must integrate Interactive Content Objects (ICOs). ICOs is the generic term for interactive tools which do not restrict learners to be passive recipients, but engage them with material that is responsive to their actions. Saul et al. illustrated in his model [25] how this results in more objective assessment findings and deeper learning, because learners can learn from mistakes and make sense from unexpected situations [25]. Since soft skills are the abilities that are learnt through practice and involvement, it becomes necessary to -

3. Incorporate interactive content objects (ICOs) as a primary technique while building the assessment system.

From a tutor’s perspective, there is a whole set of requirements as well. A tutor may want to be able to see where each individual learner stands within his/her cohort but
also to be able to realise where his/her cohort stands with respect to other cohorts. These would enable the trainer to intervene between plays or even during the play if he feels the need to. This leads to the following requirement -

4. **To develop a dashboard for the tutor that presents all trainee performances individually and collectively within the cohort, and an inter-cohort performance summary through various statistical aggregations.**

Having discussed the requirements in detail, the comprehensive objective of this research with design intricacies emerges as:

A game performance feedback dashboard will be developed for the learners and the tutors at the end of each gameplay, so that they can recognise their weaknesses and honour their strengths. The dashboard will encourage them to explore the alternatives, right or wrong, so that they really begin to reflect on their discussions. In doing so, they would naturally be encouraged to connect to their tutors to inquire on why they were marked badly for questions they believed they answered correctly. Finally, this system will possess a comparison model that enables learners to see how they did between plays and learn if are they improving.

### 3.3 Design Challenges

There were a number of challenges faced to meet the above requirements. The baseline assessment system, as seen in Figure 3.5, was examined to understand the challenges faced in learning and comprehending value from the perspectives of players. The assessment report just uses star rating (a user is scored between 1 to 5 stars in each competency after a full playthrough) as the mechanism to mark the overall performance of the player. This metric is obscure as one’s notion of the number of stars constituting a "good" score is highly varied. Further, an overall score/rating has been overlooked which is usually a player’s key takeaway from any assessment feedback system. Thus, arises the first challenge of this research -

1. **To develop an arithmetic scoring mechanism, providing an overall score and a score for each competency.**

While building this arithmetic system, it was realised that giving numerical marks will also not help the learners really distinguish between how well they performed in any answer mainly because of the absence of a metric informing the user of what score is good enough and what is not.

The in-built evaluation system looked at each learning scene as of identical impor-
tance and relevance. This design although recognised internationally, it is arguable as different organisations have different methods of training. Different people may find different situations difficult or easy. Analysis revealed there are game situations which players found relatively easier than others. Thus, the research argues that giving absolute scores in all scenes with complete disregard to the difficulties of scenes is worth contemplating. For instance, scoring 5 stars in a question which is usually answered correctly by players may be less impressive than scoring 3 stars in a question which is answered poorly by almost every player. Since the game rubric is set and agreed upon internationally and is right, what was possible within the scope of this project was to give them a feeling of how well the learners performed in a question based on other players. This is where a new way of scoring was evolved, a way that is graphed but doesn’t change the actual score.

2. **The research proposes to build an adaptive scoring mechanism that considers the scene difficulties to inform the player on his/her ranking in the overall unit of players.**

This system will help the players zero-in on the places they are differentiating themselves by performing better or worse. This generates an inner challenge -

3. **To find a way of generating scores that actually differentiate performance on a per-scene basis.**

Since, the game rubric is set and agreed upon internationally, this challenge required the generation and development of an approach that would allow differentiation between a gameplayer who is not just scoring highly, but scoring highly in scenarios where other people aren’t scoring that well. In order to do this, a weighting was generated based on how the performance in a scene collectively was.

In Table 3.1 “Collective Percentage” stands for the percentage of mean of all scores of a question divided by the maximum possible score i.e. 5. This will give the percentage success rate of a scene in a competency.

\[
\text{CollectivePercentage} = \sum_{i=1}^{n} \frac{\text{score}_i}{\max} \times 100
\]  

(1)

where \( \text{score}_i \) = achieved score, \( \max \) = maximum possible score
Table 3.1: The approach to adding weights based on the users’ data

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Collective Percentage</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>90% - 100%</td>
<td>0.1</td>
</tr>
<tr>
<td>2.</td>
<td>80% - 90%</td>
<td>0.2</td>
</tr>
<tr>
<td>3.</td>
<td>70% - 80%</td>
<td>0.3</td>
</tr>
<tr>
<td>4.</td>
<td>60% - 70%</td>
<td>0.4</td>
</tr>
<tr>
<td>5.</td>
<td>50% - 60%</td>
<td>0.5</td>
</tr>
<tr>
<td>6.</td>
<td>40% - 50%</td>
<td>0.6</td>
</tr>
<tr>
<td>7.</td>
<td>30% - 40%</td>
<td>0.7</td>
</tr>
<tr>
<td>8.</td>
<td>20% - 30%</td>
<td>0.8</td>
</tr>
<tr>
<td>9.</td>
<td>10% - 20%</td>
<td>0.9</td>
</tr>
<tr>
<td>10.</td>
<td>0% - 10%</td>
<td>1.0</td>
</tr>
</tbody>
</table>

If a collective percentage of a competency in a scene was greater than 90%, then it is considered an easy scene and given a weight of 0.1 on a scale of 0-1. Similarly, if the collective percentage of another scene is between 80-90%, then the weight attached to it would be a notch higher than the previous weight, i.e. 0.2. Table 3.1 explains in a tabular form as to what weights were attached to a scene depending on high or low success rates, on a per-competency basis.

Lastly, the current assessment system handles the learning objectives at a coarse level. There has been no development at any sub-objective. There is a total of 60 sub-objectives from the 3 competencies - 20 for each - broadly categorised into 2 sections, 10 in knowledge and 10 in performance. As mentioned earlier, player interaction with their assessment results in a deeper learning [25], a detailed representation of the objectives is required for players to learn from their mistakes and understand what all skills were being tested in any particular scenario. This thesis has refined down each of the coarse competencies into the sub-objectives to provide a comprehensively deeper analysis to the learner.

4. To present learners with a detailed learning objective that better informs the player on the skills being tested apart from the three main terms: communication skills, gender awareness skill, culture awareness skill

### 3.4 Technologies used

This project is purely built on a Python framework to build charts and analyse data. Python is an easy-to-use language that has decent library availability and great community participation. The main platforms for development are the Jupyter Notebook App—a server-client application that allows editing and running notebook documents via a web browser and Sublime text—shareware cross-platform source code
editor with a Python application programming interface [26]. The main libraries used for the following stages of data analysis include-

3.4.1 Data Preparation

This project uses pandas as it’s data preparation library. pandas has a wide range of capabilities for input/output data formats, like Excel, CSV, Python/NumPy, HTML, SQL and more. Furthermore, pandas has powerful querying possibilities, statistic calculations and basic visualisations. It has a rich documentation, but a bit confusing syntax, which is often pointed out as its most significant flaw[27].

NumPy is a package that defines a multi-dimensional array object and associated fast math functions that operate on it. It also provides simple routines for linear algebra and sophisticated random-number generation. the main advantage of NumPy in the project (and generally) is its ability to perform efficient vectorized computing and broadcasting over n-dimensional arrays.

3.4.2 Data Visualisation

Visualisation as a means to inform and improve the learner’s behaviour is the main focus of this project. The best-fit visualisation tool with support of all standard graphing was chosen. Plotly is a collaborative, web-based graphing and analytics platform. It offers a Python sandbox (NumPy supported), data grid, and GUI for analysis and styling graphs. Jupyter is a convenient and well designed web application, to produce Plotly charts and perform some aggregations of data. Its main limitation is a relatively unintuitive syntax, making it difficult to learn. However, the flaw is compensated with rich documentation providing a lot of examples.

3.4.3 Deployment

Software deployment is required to make the app available for use by others. Dash is used to create this web-based application having the ability to integrate Plotly graphs into webpages. Dash is an Open Source Python library for creating reactive, Web-based applications. Dash applications are web servers running Flask and communicating JSON packets over HTTP requests. Dash’s frontend renders components using React.js, the Javascript user-interface library written and maintained by Facebook.

Once the web application is ready, a cloud platform as a service (PaaS), Heroku is used for deployment. The Heroku cloud service platform is based on a managed container (called dynos within the Heroku paradigm) system with integrated data services and a powerful ecosystem for deploying and running modern applications. It has features
for a developer to build, run and scale applications in a similar manner across most languages. Its easy to use application development and deployment has made it the top choice of many development projects. Since the Heroku platform manages hardware and servers, developers that use Heroku are able to focus on perfecting their apps — and not the infrastructure that supports them.
4 Method and Implementation

This chapter gives a comprehensive discussion on the tools, techniques and methodology applied throughout the study. It intends to inform about the various steps undertaken at each stage, from the start of this research, until its completion.

4.1 Software development lifecycle

The development cycle followed in this research to create a performance dashboard is depicted in Figure 4.1. This is a 6 stage process undertaken to produce a live reporting dashboard that reports on the player’s performance, educate her/him on the learning objectives of the game and intuitively motivate the player to build on their skills through practice and progress tracking. Each stage will be explained in the following sections.

![Software development life cycle (SDLC) of the game assessment system](image)

Figure 4.1: Software development life cycle (SDLC) of the game assessment system

Source: from ResearchGate [28], 2016
4.1.1 Data collection and preparation

The first and the second stages of the development lifecycle are combined as they are both very closely related.

The first stage is data collection. The initial data corpus was generated from a trial conducted at GAP Soft Skills Summit, held in January 2019. This data is dumped into the database of ADAPT Centre. The GAP game developers constructed a service-oriented-architecture, making data acquisition much simpler. Initial trials data and any successive playthroughs were retrieved from the configured API calls in a JSON format. The data in Figure 4.2 a snippet of a typical JSON data retrieved from a dump API call for a player. The different JSON objects, as seen in Figure 4.2, can be understood as the different modes of the game as discussed in Game mechanics, among other sections.

![Figure 4.2: A JSON file for a player’s gameplay](image)

Even though the data retrieved was from a properly structured database, some further data preparation was required. In order to add value to the existing data, like calculation of arithmetic scores, a pre-processing script was required. This script was built in Python and the processed data was saved to a separate database to be used while data modelling.

4.1.2 Data modelling

Data modelling is the core essence of this project. A number of visualisations have been developed for the end-user dashboard.
A. Performance ranking

Firstly is a visualisation that portrays how well the learners performed in a question based on other players. The Figure 4.3 is a screenshot of one such graph developed for the dashboard as part of this research proposal.

Below are your standings amongst fellow peacekeepers

The following is a walk-through on your performance from every scene. It shows your position relative to your colleagues' performance and not your absolute score. The red coloured bars indicate your position is behind 60% of the players, while the blue indicates you are ahead of half of the fellow players and green is an exceptional performance putting you in the top 15% of the players.

![Communication performance percentile (as Maria: Military liaison officer)](image)

Figure 4.3: A player’s ranking in communication competency throughout the scenes in the role of a female military liaison officer

This graph shows the percentile\(^1\) rank of the scores of a specific player for their answers in communication competency scenes in the role of Maria: a female military liaison officer. The graph is interactive giving a player the ability to:

1. switch between the 3 competencies - communication, culture and gender - using the radio buttons provided directly above the graph.

2. hover on the graph to see the percentile rank and the selected choice of the scene being hovered on.

3. zoom in and out.

By studying this graph, a player will be able to understand his/her performance and get an idea of their position in the game amongst their colleagues on a per scene per competency basis.

This visualisation is a combination of a standard and an inverted bar graph - the line of separation being the 50\(^{th}\) percentile mark. Scores denoted in green colour can be

\(^{1}\)refer Appendix B.1.1 for definition
considered excellent, meaning the player stands in the top 15% of the total players. This performance is a combination of the absolute score achieved in a competency in a scene with the corresponding weight attached to it as explained in Table 3.1. Thus, a green bar shows the player of not only his remarkable performance independently but also how most of the other players couldn’t perform well in that question. The blue bars indicate an average performance or a player’s standings is in the mid range of performers i.e. in the range of 40 – 85th percentile rank. These scenes, however, do not necessarily indicate that the player achieved an average score but that these scenes could be straightforward enough for a majority to score really well in. While the ones in red indicate a below average score informing that a player is behind at least 60% of the other game players. The red bars can be a great introspection point for the player to dig deeper into the dashboard and learn why he/she did not perform very well in those questions. A total of 6 graphs like Figure 4.3 are available on the dashboard - 2 for each competency (Communication, Gender and Culture) where the first one is for the role of a military liaison officer while the second is for the role of the police liaison officer. The graph for the second role-play i.e as Adam: a police liaison officer is presented right below the first graph on the dashboard. These two graphs, positioned one above the other, give a clear comparison of the performance in the two roles. Both the graphs are connected to the same set of radio buttons, meaning they collectively update when a competency is changed between communication, culture and gender.

B. Inter cohort performances

This visualisation is for the tutor dashboard that presents aggregated results of a cohort in comparison with other cohorts.
Figure 4.4: Inter-Cohort performances

The figure 4.4 represents a grouped bar graph. The purpose of this visualisation is to inform cohort leaders to analyse their team’s performance relative to the other cohorts. A number of statistical aggregations are provided to give tutors the ability to get a 360° analysis of the team. This interactive graphing has the ability to:

1. navigate between statistical averages of mean\(^2\), median\(^3\) and max (top performer) through radio buttons.
2. select any number of cohorts for comparison by using toggle tokens provided on the top-left corner.
3. filter competencies based on need by double clicking the required legends on the right.

The three competencies (Communication, Culture and Gender) are grouped together per cohort. The x-axis denotes the different cohorts selected by the tutor for comparison. This graph not only provides an inter-cohort performance analysis but is also a unique way to compare the different skills within the cohort. For e.g. from Figure 4.4 it can be seen that every cohort is the most skilled in communication, followed by culture and the least in gender.

\(^2\)refer Appendix B.1.2 for definition
\(^3\)refer Appendix B.1.3 for definition
C. Performance summary

The Figure 4.5 portrays some of the most fundamental arithmetic data available. The top of the figure shows the total achieved score in percentage. The matrix shows achieved, maximum and percentage scores for each competency. The calculation of each of these scores (and the total score) is done using absolute scores as defined in the original game rubric. Each player is assigned a unique game identifier for each full gameplay by GAP. To ensure security, anonymity and uniqueness, this identifier is of five words and is only available to the user themself and the tutor. The identifier is used to search for and load the dashboard for the user by entering it in a text box provided as seen in the Figure 4.5. In this case, the identifier is "boundless tinderbox morphine droop blanching".

![You achieved a score of: 80.57%](image)

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>COMMUNICATION</th>
<th>GENDER</th>
<th>CULTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achieved Score</td>
<td>149</td>
<td>62</td>
<td>118</td>
</tr>
<tr>
<td>Max Possible Score</td>
<td>184</td>
<td>76</td>
<td>149</td>
</tr>
<tr>
<td>Percentage</td>
<td>81</td>
<td>81</td>
<td>79</td>
</tr>
</tbody>
</table>

Figure 4.5: Arithmetic scores of a game play

D. Cohort leaderboard

The GAP game is developed and is meant for improvement in soft skills for consequences in the real world, but it is a game nevertheless. Therefore, it was only fair to provide a leaderboard to show who the highest (and lowest) performing players are and their scores. It also answers the customary question asked after each test - "Who scored the highest?".
This metric is however only for tutors so that they are able to learn and read their team better. This leaderboard is intended for the tutor to discern the under-performing personnel and assist them in becoming better by identifying their weaknesses and encouraging them to learn from their mistakes.

E. Performance comparison

Returning players are the most crucial players for the dashboard as they are already motivated and inclined to learn. Thus, a proper metric to inform these users on their learning and improvement is essential. The chosen form of performance comparison is shown in Figure 4.7. It is a kernel density estimate\(^4\) that shows the density of scores (the possible scores are: 5 (very good), 3 (good), 1 (poor)) achieved throughout the game. Thus a decreasing curve from left to right is a poor performance where the density of 1’s are highest, while an increasing curve with a maximum density at 5 is an above average performance. The comparison between the two trials can be seen with the density scores at 1, 3 and 5. For e.g. the user in Figure 4.7 has a fairly straight performance in attempt 1 with a little higher number of 5 scores. Attempt 2 however shows an increasing curve with considerably lower densities for 1 and 3, and a sharp escalation in 5. Thus, the user can conclude an improved performance through the inclination level of the increasing curve.

---

\(^4\)Refer Appendix B.1.4 for definition
4.1.3 Evaluation

Evaluating the usefulness of a data visualisation is subjective, for it is based on an assessment of the needs and values of others, but that does not diminish the relevance of this criterion and the attempts to measure it [29]. The data visualisation effectiveness profile that was proposed by Stephen consists of seven criteria, which fall into two general categories: criteria that address the degree to which a visualisation is informative (i.e., produces understanding) and criteria that address the degree to which it is emotive (i.e., produces a useful emotional response). The models presented in section 4.1.2 can be categorised effective and are hence displayed on the dashboard. Many of the graphs which failed to effectively bring out the intended value remain as exploratory analysis and visualisations in the project. A preview of a few exploratory visualisations include -
Figure 4.8: Chapter performance per competency

Figure 4.9: A waterfall chart to visualise learning rate
4.1.4 Deployment

The objective of deployment phase is to make the developed software operational in a live environment. This project is deployed locally, meaning an active version of the dashboard is operational on the author’s machine. A fully deployed version of the dashboard is pending of now as the dashboard is still in a testing phase and due to the recent outbreak of Covid-19, it has not been possible to get acceptance by the game development authorities.

4.1.5 Monitor

The performance of learning models and produced knowledge graphs is monitored continuously through reporting tools and usage logs. Based on the feedback during the monitoring phase, development process will continue to evolve to ensure that information shifts are handled and newly emerging knowledge patterns are incorporated [28].

4.2 Workflow

There are 2 main workflows:

1. between the game and the assessment engine.
2. between the assessment system and the different users.
4.2.1 Workflow between the game and the assessment engine

The workflow between the game and the assessment system has been scripted to work fully automatically. Any player playing the game has their data synced with ADAPT Centre’s database servers. This data is then retrieved from the servers through API calls configured by a service-oriented architecture. Once the data is retrieved, an extract transform methodology is applied to process the data.

Extract, transform, load are three database functions that are combined into one tool to pull data out of one database and place it into another database [30].

**Extract** is the process of reading data from a database. In this stage, the data is collected from multiple and different types of sources.

**Transform** is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.

**Load** is the process of writing the transformed data into the target database.
Once the data is prepared in the usable format, it is fed into the graphs discussed in section 4.1 for player assessment and learning.

4.2.2 Workflow between the assessment system and different users

There are two dashboards developed for 2 different user base. One is for the personnel playing the game and looking for learning insights and the second is for the tutors who are supervisors of those personnel. Following are walkthroughs of the dashboard from the perspective of the two user base.

A. Student dashboard

The player/student after completion of the game is directed to the dashboard for results. The first landing page contains a left-sided navigation bar to switch between multiple pages of the dashboard. The content on the landing page is of the game overview page. It contains the arithmetic summary and relative ranking of the players as described in Figure 4.5 and 4.3 respectively. The second page shows the selected answers with the competency scores attached as in Figure 4.14. From this page the
Figure 4.13: Navigation bar on the left with the player’s arithmetic summary beside it.

student is able to click through an answer he/she may want to explore. This will lead him/her to the "Review all answers" page that displays all possible answers- right and wrong - with the correct answer highlighted in blue. The third page is then again the "Review all answers" as in Figure 4.15, where all scenes with their corresponding options are disclosed. The novelty here is the disclosure of all the sub-learning objectives hidden in each scene. This gives the player a deeper insight to what all parameters the scene was looking out for in the answer. Lastly, is the fourth page of the dashboard shown in Figure 4.16. It is the feedback page. This page is added to receive feedback from the players actually playing the game. The intention here is to learn from first hand users how they felt about the game and the assessment system and mostly importantly if they have any suggestions for improvement of the system.

4.2.3 Tutor dashboard

Tutors have a dashboard separate from students. This is so to keep player performances anonymous to each other. Tutors can login anytime to see the performances of their team. There are a total of 5 pages in this dashboard.

The tutor on login gets an exclusive access to aggregated results like leaderboards as seen in Figure 4.6 and inter-cohort comparisons like in Figure 4.4 on the landing page. The tutor dashboard structure is similar to that of the student dashboard with a vertical navigation bar on the left and insightful charts on the right hand side. The purpose of the second page is to investigate individual performances, where the page mirrors the student dashboard’s landing page. The tutor will have to enter a student’s player id to learn about his/her performance. The third and the fourth page are also similar to the students’ dashboard where the tutor’s third page, i.e. "Individual Answers", shows the answers chosen by an individual the tutor is learning about. The fourth page, i.e. "Review all answers", is the full metric of correct answers with the associated sub-learning objectives. Lastly, there is a feedback page for tutors as well. Even though a
Figure 4.14: Choices chosen in the gameplay with their corresponding scores are displayed.

Kumail: Hello, beautiful! What is a delicate flower like you doing so far from home?

"Smile." I'm just trying to escape the guys back in Finland.

Kumail: Hello, beautiful! What is a delicate flower like you doing so far from home?

I'm here to do my job as a peacekeeper, that's all.

Kumail: "His smile is mischievous." Tell me, are you married?

(Truth) No, I'm not married.
Figure 4.15: A screen-grab of the dashboard describing all the sub learning objectives associated with a scene and the best fit solution in it

Figure 4.16: A feedback section in the assessment dashboard for end-user opinions
A tutor might not be playing the game, their feedbacks are equally valuable as students. Tutors are mentors and know their team the best. Thus, they might be able to perceive the students’ assessment dashboards better than the student themself.
5 Evaluation

A feedback assessment system for GAP was developed to provide learners with a personal dashboard in order to increase their awareness on soft skills and track their learning progress. While feedback is a useful mechanism, instances like the one provided in Figure 3.5 are not as effective as are needed for positive student growth. The most effective type of feedback is highly personalised and relevant to the subject area being assessed. To further personalise the ‘Learner’s Centred feedback’ for GAP, a live post-gameplay dashboard has been designed and implemented.

The original intention was also to have an end-user evaluation of the usability and efficacy of the learning analytics on a live dashboard. 15 users had signed up from Trinity research community and engineering department for the testing phase. However, due to the global outbreak of coronavirus (SARS-CoV-2), the research base had to be shifted from Ireland to India and with the country’s guidelines on a nationwide lock-down, the trials could not be conducted.

One user trial however was conducted at home. Figure 4.7 is the actual end-user performance comparison graph as displayed on the game dashboard. The player showed a 0.6% increase in their performance after the first interaction with the feedback dashboard. A brief questionnaire attached in Appendix C, was prepared to judge the usability and efficacy of the dashboard from an end-user’s perspective.

Thus in this research, the evaluation of the customised feedback dashboard can be considered a ‘success’ based on the fulfilment of the following requirements:

Requirement 1 A dashboard has been chosen as a reporting mechanism for trainees to reflect upon their performance for an actionable and interactive game-based learning.

Requirement 2 The dashboard developed is live and synced with the game database for immediate feedback to the learners after every gameplay.

Requirement 3 The developed dashboard is a multi-page reporting platform that is built with aesthetically pleasing and interactive content objects like- input boxes,
buttons, toggle tokens and switches to engage the players with the feedback platform and motivate them to learn from mistakes with a positive attitude.

**Requirement 4** A separate dashboard has been built for the tutors to assess their team’s performance. This dashboard also provides the tutor with inter-cohort analysis that is intended to give tutors an idea where their team stands between other international contingents.

In addition to satisfying the requirements, a number of challenges were also overcome successfully:

**Challenge 1** A score summary is added to the front page of the dashboard that includes an overall percentage, an absolute score matrix with percentages for each competency.

**Challenge 2** In an attempt to fairly judge the difficulty level of a scene, an adaptive scoring mechanism using a weighted scale as in Equation 1 was developed to rank different scenes.

**Challenge 3** The integration of challenge 2 with absolute scores achieved in the game help distinguish learners performance on a scene difficulty basis and eventually ranks the player on it. It is one of the novel concepts addressed in this research to provide extrinsic motivation to continue playing for game-based learning.

**Challenge 4** The sub-learning objectives associated with a scene is disclosed to the player for a better understanding of the requirements of the scene and learn different perspectives.

### 5.1 Discussion

This section discusses to what extent each of the requirements were met and the challenges were overcome. It pertains to whether or not there is potential for improvement in the current scope of the research design.

**Requirement 1** focused on the creation of a reporting mechanism and the development of a web-based feedback dashboard which allowed easy accessibility. This aim has successfully been achieved.

**Requirement 2** was to provide a live feedback on performance. This ‘live’ assessment, however, is only available after the completion of gameplay. The dashboard was purposely built as an end-game feedback as the learning analytics described in the dashboard aim at providing insights on a per-scene basis and cumulative aggregations for performance results and trends. However, a further immedi-
acy of assessment through an in-game feedback is possible. This would require access to game licensing and since the project is separate to the game and as designed to be used after completion, an in-game feedback that tells the user about the correctness of their answer right after selection falls out of the scope of this project.

**Requirement 3** emphasised on the integration of Interactive Content Objects (ICOs) in the reporting platform. This has been achieved by using highly customisable graphing library; *Plotly* producing interactive, production-quality graphs. This was accompanied with a *Dash* framework allowing the dashboard to run on web browsers with a comprehensive design kit containing hundreds of interactive components including, but not limited to, text boxes, clickable events, toggle tokens and navigation bars. Also, the dashboard has been designed as a multi-page app in order to break the dashboard into several pages to make it less overwhelming to the user and present data in a very organised fashion. One ICO on creating clickable graphs for easy navigation to areas of interest was initially intended as the part of the project. This is for the performance ranking graph shown in Figure 4.3. It currently is **hoverable**\(^1\), which means that a user will be able to learn about the choice made in the scene if a cursor is placed on top of the corresponding visualisation section. But the same figure does not have the functionality to actually click on it and get navigated to the full description of the scene. Due to time constraints and unavoidable circumstances, it remains as a to-do and would be a valuable addition to the dashboard if and when it is included. Additionally, a similar suggestion as seen in Appendix C, was offered by the one user who tested the dashboard.

**Requirement 4** of the research solely was to build a separate dashboard for tutors. Such a dashboard has been implemented which matches its counterpart in interactivity and design. A number of quality visualisations have been incorporated to the dashboard and the requirement can be considered fully met.

**Challenge 1** an arithmetic result for the learners is presented using the absolute scores. This in itself is an informative metric on the performance in every competency, but an additional score could also be presented using the adapting scoring mechanism.

**Challenge 2** the adaptive scoring technique has been successfully implemented to rank different scenes. An extension to this could be the disclosure of these difficulty levels during the game. Since this was an independent study from the game, the integration of the difficulty badges within the game is proposed by

\(^{1}\)the feature of presenting a summary of an element when the user moves the mouse over it.
the author but it simply fails to lie within the current scope of the project.

**Challenge 3**  the usage of adaptive scoring technique in the performance ranking graph appears promising. The user evaluation also reveals the usability and efficacy of this visualisation. A possible improvement as suggested in requirement 2 above is the incorporation of clickable events which would enhance learning by increasing accessibility.

**Challenge 4**  there are a total of 60 sub-learning objectives distributed between the 3 competencies. These are disclosed to the learner in the dashboard to give a clearer understanding of the precise skills that the scenes are testing. It was felt that it would be beneficial if a way of assigning scores to these sub competencies would be enabled. However, that would require a lot more detail being provided by the GAP game designers regarding the sub-learning objectives. Due to the lack of such data, a deeper analysis of these sub-learning objectives, correlation between them and the achieved score and the consequent improvement in these scores is not possible at the time.
6 Conclusion

“Peacekeeping training entails much more than unarmed combat exercises, marksmanship, and obstacle courses; on the ground, the most important talent may be walking in the shoes of the native population... The quality you need most in United Nations peacekeeping is empathy.” [31, Canadian peacekeeping soldier]

Educating individuals on building effective interpersonal skills is critically important in peacekeeping training and CPPB missions. It requires a “soldier-diplomat” to stop people from fighting [32]. Gaming for Peace was developed to train soldiers in simulated peacebuilding situations as real scenarios are expensive and impractical to recreate. This research has built upon the GAP’s intention of providing a simulated training platform for soft skills learning to enhance learning through immediate and interactive feedback for game-based learning. This research also provides an immediate and interactive platform to the tutors for a comprehensive analysis of their team.

The immediacy of feedback in learning analytics raises the aspiration levels in learning tasks. This is due to the fresh retention of topic and a positive motivation to learn from mistakes. The interactivity characteristic of the feedback dashboard enables active, constructive and long-lasting learning.

The user-centric dashboard allowed learners to access their performance whenever they wished. This helped to increase student engagement and motivation. The dashboard also escalated active engagement of the user because of having interactive capabilities. Therefore, the customised dashboard can be inferred as a constructive tool for encouraging learners to use the GAP game more frequently. This was one of the principal objectives as recurrent game-plays enables students to build the necessary interpersonal skills.
6.1 Future Work

Future work will involve end-user trials for an accurate analysis on the usability and efficacy of this feedback dashboard. Since testing was limited to only one user due to reasons mentioned in Chapter 5, a much more rigorous testing phase would be the next step to establish concrete results. A fully commissioned feedback dashboard will follow through the successful completion of this user testing. This will then enable the assessment system to be functional anywhere in the world eliminating roadblocks like a global pandemic disrupting the progress of the research. Further, the integration of this system with the game will follow once the user testing results are approved by the GAP game editorial board.

6.2 Future Scope

A longer-term vision would be the implementation equivalent to Kirkpatrick’s level 4 - Organisational performance; after an organisation’s entire staff has used GAP and the feedback dashboard a certain number of times and consequently participated in real peacekeeping missions, it would become possible to focus on and measure difference in job behaviour and benefits to the society.
Bibliography


A Appendix

A.1 Learning Objectives

A.1.1 Communication

As previously mentioned effective communicators are essential to ensure the delivery of the quality and range of tasks expected in peacekeeping missions. Providing the direction and vision for an equitable, well-managed and values-driven service delivered by civilian, military and police actors is key to ensuring organisational, national, mandated and community needs are met as far as is possible and practicable. The purpose of this topic is to enable learners to develop an understanding of common communication and listening theories and styles, and how these compare and contrast in different organisations involved in missions, and to understand key principles such as confidentiality, building trust, respect and demonstrating understanding and compassion to achieve mission vision and objectives. It will provide opportunities for learners to consider and explore practical examples of de-escalation through communication as well as leading through communication in action in a peacekeeping service context. In order to engage with these topics effectively, learners must be:

1. reflective learners
2. team workers
3. self-managers.

Learners must demonstrate (knowledge): [KEY=CMK(1)...]

1. key theories and models of communication, including active listening and trust building [applied] within personal, internal, and mission communications, and how they apply to military, police and civilian organisations

2. what is meant by:

   (a) gender
(b) sex

(c) masculinities and femininities

(d) sexualities, heteronormative, transgender, homophobia

(e) sexism and how it manifests itself within organisations

3. how to source gender related information including UNSC 1325 and other relevant directives, feminist theories and gender mainstreaming regulations

4. a deep understanding the benefits and disadvantages of differing communication methods and their effects on relationships and information sharing

5. the importance of non-verbal communication, such as body language, and how different cultures use and interpret body language in different ways

6. the role of effective communication in reducing and de-escalating conflict, including the type of constructive behaviour you can take to defuse situations including body language, spoken language, posture, emblems such as illustrators, affect displays, regulators, adaptors and para-language

7. when it is recommended to use an interpreter who can convey the message with the needed level of accuracy and precision, and methods for ways of working with interpreters’/language assistants to achieve mission goals

8. how addressing the language barriers and working with interpreters impacts on communication

9. procedures and requirements for formal information sharing between relevant organisations including the restrictions on the disclosure of sensitive information

10. how partner organisations are organised including:

   (a) their broad structures

   (b) methods of communication

   (c) decision making processes

11. the importance of keeping official and private information separate, and how to keep all data secure

12. the role of emotional intelligence in communication models including recognising own feelings

13. how to establish and maintain effective communication in missions, with a range of actors, which ensures integrity, respect, and transparency
Learners must be able to demonstrate (performance): [KEY=CMP(1)…]

1. trust, rapport and communicate in a polite, respectful, ethical, timely, patient and culturally-appropriate manner

2. effective communications methods, including written, verbal and non-verbal, contextualised to the situation in order to:
   
   (a) manage conflicts

   (b) establish shared understanding with the range of actors

   (c) achieve mission objectives

3. source and share information using common language and terms with other organisations through an on-going culture of dialogue and co-operation in line with mission mandate and organisational processes

4. recognize, and act on, discrimination based on culture, gender or sexuality

5. work with an interpreter during risky negotiations, highly complex meetings or when detailed and sensitive information is being used

6. keep secure records of expectations, conversations and agreed actions

A.1.2 Cultural Awareness

Engaging communities, both internal communities of others mission actors and external communities of the local population, in determining their own needs and possible solutions to community issues helps to promote ownership and cooperation. Cultural distinctions and showing awareness of these is a key element to demonstrating empathy, emotional intelligence, and trust all aspects which ensure missions are appropriate, relevant and targeted to meet both mandate and community needs. The purpose of this topic is to introduce learners to the way civilian, military and police actors engage with diverse communities, and to respond to changing needs. It will also enable learners to explore the way in which culturally distinct and aware partnerships work between communities and actors to deliver citizen-centred responses which are equitable, universal, empowering and proactive.

Learners must demonstrate (knowledge): [KEY=CLK(1)…]

1. how to analyse the intersection of diverse organisational and national cultures related to your mission

2. the power dynamics within internal and external communities especially due to gender, caste, class, ethnicity, and disability which may make it difficult to hear
some voices/opinions

3. the range of tools used in encounters where diversity, tensions and conflicts can be expected to arise and a clash of cultures is often inevitable

4. what is meant by:
   (a) gender
   (b) sex
   (c) masculinities and femininities
   (d) sexualities, heteronormative, transgender, homophobia
   (e) cultural discrimination and how it manifests itself within organisations

5. how to source culture awareness related information including UNDPKO standards and other relevant directives, theories and cultural competency regulations

6. the importance of recognising that missions bring heterogenous personnel into contact with a range of in-mission organisations and local people who often draw upon cultural background different from those of own organisation and staff

7. systems of building trust, and the importance of doing so for success of the mission

8. the types of power relations within different contexts are identified and discussed in terms of the impact on equity and respect based on culture or gender

9. how culture, gender and beliefs can affect what is perceived as ‘acceptable’ and ‘non-acceptable’ behaviour (such as it may be seen to be more acceptable for men to be assertive than women)

10. how the worker’s own culture, gender and beliefs affect the way that they view the behaviour of others and why it is important to recognise and challenge this

11. sources of information that provide realistic overview of the full range of in-mission attitudes including that of other actors and organisations as well as local attitudes

12. range of engagement styles to use with different actors

13. Analyze own understanding of one’s own cultural background because your nationality/country of origin/religion has shaped this and others will have different socialization into culture roles
14. Knowledge of the positive differences culture can make in an international mission with diverse cultures in organizations and in local population

15. Analyze and understand one’s own and others organizational culture and organizational practices.

Learners must be able to demonstrate (performance): [KEY=CLK(1)...]

1. build bridges of trust between yourself, your organisation, other mission organisations and the host community, including:
   
   (a) review anthropological and cultural guides about the different peoples you will be interacting with beforehand

   (b) analyse own cultural background because your nationality and country of origin may have a historical footprint in colonialism and occupation

   (c) learn about other culture’s customs and history, whilst acquiring a deeper understanding of your own

   (d) learn some basic phrases in the range of languages likely to be used on-mission to reflect interest in the cultures of other actors and your respect for individuals

2. describe how cultural differences and different kinds of diversity might be evident in the mission environment between organisations, and in the host country, and how to show respect for this in a multicultural peacekeeping environment

3. recognize, and act on, discrimination based on culture

4. take action to maintain calmness and safety in the working environment which values individuals with as little restriction of action as possible to encourage meaningful interactions

5. engage with relevant leaders, powerbrokers and personalities to foster good relations between organisations and with local populations

6. work with other organisations to strengthen capacity to respond to community-identified protection needs

7. analyse key actors directly or indirectly involved or who have influence over a particular problem including the role of power relations

8. Ability to use the positive differences culture can make in an international mission with diverse cultures in organizations and in local population
9. Ability to recognize and act upon one’s own and other organizations’ cultures and practices.

A.1.3 Gender Awareness

Please note: this topic is an accumulation of the gender aspects embedded with the other 6 GAP soft skills topics. If using this as a standalone topic all aspects must be completed to be valid. If being completed with other topics it is advisable to conduct a quick mapping across this topic and the others to avoid duplication.

In the UN DPKO/DFS 2014-2018 Gender Forward Looking Strategy, the UN states that, ‘Equality is fundamental to lasting peace and security. It is therefore critical that all members of society have equal access to opportunities, resources, services, protection, decision-making and other basic rights’. The purpose of this topic is to enable learners from civilian, military and police sectors to develop an understanding of gender issues within their daily practice, and how to embed this within all aspects of their work. It will provide practical insights and enable learners to rearticulate the gender architecture in peacekeeping operations in light of a changed and further evolving landscape.

Learners must demonstrate (knowledge): [KEY=GK(1)...

1. knowledge of theories of personality and behaviours
2. theories of identity and self esteem
3. theories/good practice relating to the use and misuse of power and discrimination
4. what is meant by:
   (a) gender
   (b) sex
   (c) masculinities and femininities
   (d) sexualities, heteronormative, transgender, homophobia
   (e) sexism and how it manifests itself within organisations
5. how to source gender related information including UNSC 1325 and other relevant directives, feminist theories and gender mainstreaming regulations
6. the importance of non-verbal communication, such as body language, and how different cultures and genders use and interpret body language in different ways
7. how culture, gender and beliefs can affect what is perceived as ‘acceptable’ and ‘non-acceptable’ behaviour (such as it may be seen to be more acceptable for men to be assertive than women)

8. how the worker’s own culture, gender and beliefs affect the way that they view the behaviour of others and why it is important to recognise and challenge this

9. the power dynamics within internal and external communities including inter-organisation especially due to gender, caste, class, ethnicity, and disability which may make it difficult to hear some voices/opinions

10. concept of power relations

11. role and impact of traditional practices on self, own organisation, other organisations and wider society is explained in terms of how it perpetuates gender stereotyping and inequality

12. the types of power relations within different contexts are identified and discussed in terms of the impact on gender inequality

13. power relations include but are not limited to class, race and sexism

14. Analyse own understanding of one’s own gender role because your nationality/country of origin/religion has shaped this and others will have different socialisation into gender roles

15. Knowledge of the positive differences gender can make in an international mission with diverse cultures in organisations and in local population domestic violence, checkpoints,

16. Awareness of discriminatory practices based on gender in one’s own or other organisation.

17. Awareness of discriminatory practices based on gender among local population.

**Learners must be able to demonstrate (performance): [KEY=GP(1)...]**

1. demonstrate the capacity to be empathetic and understand another person’s experience from their perspective

2. build rapport and communicate in a professional, timely, patient and appropriate manner

3. recognise, and act on, discrimination based on gender or sexuality

4. demonstrate commitment to humanitarian principles
5. describe how cultural differences and different kinds of diversity might be evident in the mission environment within own organisation and other organisations, and in the host country

6. analyse the activities, motives and interests of actors on the problem, as well as the relationship between actors

7. analyse the role of institutions in shaping the attitudes and perceptions of women, men, girls and boys

8. maintain fairness for all parties involved in the process

9. Take actions based in daily interaction based on awareness of how gender roles differ for men and women, and among men and among women.

10. Ability to use the positive differences gender can make in an international mission with diverse cultures in organisations and in local population

11. Challenge masculine beliefs and behaviours that oppress women in the peacekeeping organisations.

12. Challenge masculine beliefs and behaviours that oppress women in the local population.
Appendix

B.1 Basic Statistical methods

B.1.1 Percentile

A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations falls. For example, the 20\textsuperscript{th} percentile is the value below which 20\% of the observations may be found. Similarly, 80\% of the observations are found above the 20\textsuperscript{th} percentile.

The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. For example, if a score is at the 86\textsuperscript{th} percentile, where 86 is the percentile rank, it is equal to the value below which 86\% of the observations may be found (carefully contrast with in the 86\textsuperscript{th} percentile, which means the score is at or below the value below which 86\% of the observations may be found—every score is in the 100\textsuperscript{th} percentile). The 25\textsuperscript{th} percentile is also known as the first quartile ($Q_1$), the 50\textsuperscript{th} percentile as the median or second quartile ($Q_2$), and the 75\textsuperscript{th} percentile as the third quartile ($Q_3$). In general, percentiles and quartiles are specific types of quantiles [33].

B.1.2 Mean

For a data set, the arithmetic mean, also called the mathematical expectation or average, is the central value of a discrete set of numbers: specifically, the sum of the values divided by the number of values. The arithmetic mean of a set of numbers $x_1, x_2, \ldots, x_n$ is typically denoted by $\bar{x}$[34].

The arithmetic mean (or simply mean) of a sample $x_1, x_2, \ldots, x_n$, usually denoted by $\bar{x}$, is the sum of the sampled values divided by the number of items in the sample-

$$\bar{x} = \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right) = \frac{x_1 + x_2 + ... + x_n}{n} \quad (1)$$
B.1.3 Median

In statistics and probability theory, the median is the value separating the higher half from the lower half of a data sample, a population or a probability distribution. For a data set, it may be thought of as the "middle" value. For example, the basic advantage of the median in describing data compared to the mean (often simply described as the "average") is that it is not skewed so much by a small proportion of extremely large or small values, and so it may give a better idea of a "typical" value.[35]

B.1.4 Kernel Density Estimate

In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample [36].
C Appendix

Feedback of the dashboard

How would you rate the INTUITIVENESS of the feedback dashboard?

1  2  3  4  5
(lowest)  ○  ○  ○  ○  (highest)

How would you rate the DESIGN of the feedback dashboard?

1  2  3  4  5
(lowest)  ○  ○  ○  (highest)

How would you rate the QUALITY of CONTENT of the feedback dashboard?

1  2  3  4  5
(lowest)  ○  ○  ○  ○  (highest)
How would you describe the QUANTITY of CONTENT of the feedback dashboard?

- Should be a lot more
- Should be a little more
- Perfect quantity
- Should be a little less
- Should be a lot less

Did the feedback dashboard motivate you to improve your score in the GAP game?

- Demotivated me a lot
- Somewhat demotivated me
- Had no affect in this regard
- Somewhat motivated me
- Motivated me a lot

There are a number of graphs and visualizations in the feedback dashboard along with their respective descriptions. Do you think they are easy to understand?

1 2 3 4 5

(very difficult to understand)  o  o  o  o  (very easy to understand)
Would you like us to add an option to compare your result with a particular colleague (with their permission) using the tools provided in the dashboard?

- [ ] Yes, it would be helpful.
- [x] No, it is not necessary.

Please let us know if you have any suggestions for us to improve our dashboard.

Wherever red bars are shown, it should be accompanied with a short explanation.

This content is neither created nor endorsed by Google.