A LOGS BASED VERIFICATION TOOL OF SERIOUS GAMES TARGETING AUTISTIC CHILDREN

Lappeenranta-Lahti University of Technology, LUT

Master of Science (Tech.) in Software Engineers for Green Deal, Master's thesis

2023

Arini Nur Rohmah

Examiners: Professor Jari Porras (LUT University)

Professor Henry Muccini (University of L'Aquila)

Professor Patricia Lago (Vrije Universiteit Amsterdam)

SE4GD
Software Engineers for Green Deal

With the support of the
Erasmus+ Programme
of the European Union

This thesis has been accepted by partner institutions of the consortium (619839-EPP-1-2020-1-FI-EPPKA1-JMD-MOB).

Successful defence of this thesis is obligatory for graduation with the following national diplomas:

- Master of Computer Science (University of L'Aquila)
- Master of Science in Technology (LUT University)
- Master of Computer Science (Vrije Universiteit Amsterdam)

ABSTRACT

Lappeenranta-Lahti University of Technology LUT

LUT School of Engineering Science

Software Engineering

Arini Nur Rohmah

A Logs Based Verification Tool of Serious Games Targeting Autistic Children

Master's thesis

2023

94 pages, 47 figures, 17 tables and 4 appendices

Supervisor: Assistant Professor Nelly Condori-Fernandez (CiTIUS-University of Santiago de Compostela)
Tutor : Associated Professor Miguel Luaces (University of A Coruña)
Examiners: Professor Jari Porras (LUT University)
Professor Henry Muccini (University of L'Aquila)
Professor Patricia Lago (Vrije Universiteit Amsterdam)

Keywords: Verification tool · Game verification tool · Logs based verification tool · Serious game · Inclusive game · Game logs · Player logs· Stress · Autism · Autistic children.

Serious games are referred as entertaining tools with a purpose of education, where players cultivate their knowledge and practice their skills through overcoming numerous hindrances during gaming. The correct implementation of serious games is extremely important to ensure a high level of adoption of the competencies by the game users, which can be even more challenging if we target autistic children. Understanding the actual needs of Children with Autism Spectrum Disorder (ASD) and supporting their learning activity through serious games can be a burden especially for game developers. This master thesis presents logs based verification tool, named Game verification Tool (GVT), to assist game developers identify potential scenes of the serious games that might be causing any negative emotion on the autistic children. Our approach focuses on monitoring negative emotions, like stress, derived from physiological data. GVT architecture follows a client sever model, which enables to merge the game logs (hosted at the client side) with the emotional data logs gathered via Bluetooth. The data analysis and visualisation which carried out at the server side, should provide relevant information to support developers to enhance the quality of serious games targeting autistic people.

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, the Most Merciful.

First of all, I would love to say thank you so much for my family, my Mom and Dad who always love, support, and encourage me to do whatever it takes until I come this far. Thank you for all the never-ending prayers you always ask for your children, I believe those are the main fuel to take off my rocket. Not to forget, my Sister, thank you for being my perfect role model who I always look up to. I really treasure our siblings love-hate relationships, if I could replace you with anything, I wont.

Thank you for my lovely, kind, amazing supervisor, Nelly Condori Fernandez, who has been full of love, care, and patience in teaching and showing me the way since the day one I met her. You are the best supervisor I could ask for!

For Miguel Luaces, thank you for being a nice tutor whose the advice I've always waited for. Thank you to be there for me to have a monthly meeting even when you have to wait a bit for my usual lateness.

Thank you for Erasmus Mundus SE4GD, to sponsor me doing my master study in European Countries. Thank you to allow me to pursue a master degree at the University of L'Aquila, Lappeenranta-Lahti University of Technology, and Vrije Universiteit Amsterdam. This is my best achievement so far in my life.

Big thank you for Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), to provide me with the excellence place to work on my master thesis. The best, comfortable workplace in Santiago surrounds by great people. It is my pleasure to be part of the team in the last six months.

For my emo, random, and supportive friend: Aji Ari Adam, thank you for keeping up with all my stupidity, weirdness, and sudden talks and requests anytime I need.

Last but not least, thank you for all of my other supportive and loving friends who are too many names for me to write down here. Without your love, support, and pray, I would not be as I am now.

Arini Nur Rohmah

Santiago de Compostela, July 2023

SYMBOLS AND ABBREVIATIONS

Symbols

Avg Stress Scene	average stress in scene
E0	event-0
E1	event-1
E2	event-2
En	event-n
t	time
ta	time accumulation
ts	time spent
tScene	total time in scene

Abbreviations

AHP_HeGES Analytic Hierarchy Process Based Holistic Online Evaluation System for Educational Computer Game

ANS Autonomic Nervous System

ASD Autism Spectrum Disorder

BT Bluetooth

CiTIUS Centro Singular de Investigación en Tecnoloxías Intelixentes

CORGIS Challenge Originating from Recent Gameplay Interaction Scale

CSL Computerized Speech Lab

DBP Diastolic Blood Pressure

DES Differential Emotions Scale

ED Emotion Detector

EDA Electrodermal Activity

EEG Electroencephalograph

EMG Electromyography

FACS Facial Action Coding System

FPA Finger Pulse Amplitude

GAME Game As a Measurement Environment

GEQ Game Experience Questionnaire

GSR Galvanic Skin Response

GUI Graphical User Interface

GUR Games User Research

GVT Game verification Tool

HBA Heart Beat Amplitude

HE Heuristic Evaluation

HEP Heuristic Evaluation for Playability

HR Heart Rate

HRV Heart Rate Variability

MAACL-R R-Multiple Affect Adjective Checklist

MACL Mood Adjective Checklist

MEEGA+ Systematic Model to Evaluate Educational Games

PANAS-X Expanded Form of the Positive and Negative Affect Schedule

PENS Player Experience of Need Satisfaction

PEP Pre-Ejection Period

PHE Playability Heuristic Evaluation

PHEG Playability Heuristic for Educational Game

POMS Profile of Mood States

PTE Pulse Transmission time to the Ear

PTF Pulse Transmission times to the Finger

SBP Systolic Blood Pressure

SCL Skin Conductance Level

SDT Self-determination Theory

SG Serious Games

SPGQ Social Presence in Gaming Questionnaire

UI User Interface

UsaECG Usability of Educational Computer Games

V&V Verification and Validation

VGDS Video Game Demand Scale

VT Verification Tool

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Information technology is a well-known field that can combine multidisciplinary works. We can see how various innovations were created from combining the fields of computer science with other applied science can assist people. The advancement of software engineering combined with the health sector can help people with special needs, allowing to promote inclusiveness. In this project, author's main focus is to assist game developers in identifying elements within their games that have the potential to trigger high levels of stress in players with Autism Spectrum Disorder (ASD).

1.1 Problem

One of the main common trait of children with ASD is a lack of stress management (White et al. 2009). The absence of stress management encompasses the following characteristics: quick boredom, outbursts of anger, and emotional meltdowns, all of which contribute to their aggressive behaviors (Overskeid 2016; White et al. 2009). Due to the impact of these relationships, there is a significant likelihood that such aggressive behaviors occur when the children engage in educational activities, particularly at school (Machalicek et al. 2007). Meanwhile, ensuring emotional stability serves as an effective method for preventing aggressive behavior in children with autism (Berkovits et al. 2016).

Utilizing Serious Games (SG) is considered one of the most effective approaches to engage children with ASD in enjoyable educational activities (Tsikinas and Xinogalos 2018). However, SG can potentially impact the stress levels since it may elevate mental workload in correlation with learning effects (Cowley et al. 2014; Zhonggen 2019) especially for children with mental disorders. Therefore, in order to minimize potential negative outcomes when children with ASD engage in SG, it would be advantageous to have a tool that can assess and analyze the suitability of SG for autistic children that taking into account these specific aspects.

Numerous guidelines, case studies, and verification tools have been developed for both games and SG, with the purpose of evaluating, identifying, validating, and analyzing various game design elements . While many of these resources primarily focus on improving game design to enhance player satisfaction, those of that may also be utilized to optimize the game experience for specific user groups such as autistic children. Additionally, there was an interesting case study related to design guidelines for serious games that are targeting autistic people (Tsikinas and Xinogalos 2019). However, there has been a limited emphasis on the develop-

ment of Verification Tool (VT) specifically aimed at examining the feasibility of games for the targeted players with the goal of maintaining their emotional stability. Creating a dedicated VT for this purpose can greatly assist game developers in identifying elements within their games that may induce high stress levels in autistic players. This, in turn, will enable developers to design serious games that minimize stress and create a more comfortable and enjoyable experience in learning activity for autistic players.

1.2 Goal and Research Questions

In this master thesis work, the development of a Game Verification Tool (GVT) aims at providing support to enhance the design of serious games for autistic children was proposed. This tool will guide game developers in analyzing game elements that may trigger high stress levels. The GVT integrates game player logs and emotion logs, which represent the player's stress level. The GVT will be connected to an emotion detector app on a mobile device through web client-server. Players will wear a wearable sensor that generates real-time data which then be sent to the mobile app Emotion Detector (ED) by Bluetooth (BT). By establishing a web client-server connection between the GVT and the mobile device, the GVT will retrieve data from the app and analyze it.

In this study, the work is served the purpose to answer the following **research questions**:

RQ1 What technologies or methods are used for games verification tools?

Rational: Different methods serve different purposes. The main emphasis of this research question is to collect and cluster the existing methods used to verify games into a scheme accordingly to understand the differences between them.

RQ1.1 Is the method applicable for serious games?

Rational: Game and serious game are intended to serve different purposes. However, they may have something in common. The main emphasis of this sub-research question is to analyze to what extent methods used to verify games can be used to verify serious games.

RQ1.2 Is the method applicable for serious games targeting autistic children?

Rational: To encourage inclusiveness, there are serious games targeting for autistic people. Since it is targeting people with special needs, the structure or element of the serious games might be different compared to those of that targeting for non-autistic people. The primary focus of this sub-research question is to analyze to what extent methods used to verify serious games can be used to verify those of that targeting autistic people.

RQ2 Can player logs and emotion logs be used for evaluating serious games for autistic children?

Rational: Player logs is generated by games to record what happens inside the game during game play. Meanwhile, emotion logs is generated by emotion detector, which is considered external device. The primary focus of this research question is to find a possible way to use those two separate logs to evaluate serious games for autistic children.

RQ2.1 How logs should be structured to be eligible for evaluating serious games for autistic children?

Rational: How game generates its player logs is entirely the discretion of the game developer. However, there should be a template that must be followed by game developers to harmonize the logs structure, so that all serious games can be evaluated equally, especially those of that targeting autistic children.

RQ2.2 How to integrate game players logs and player's stress level in order to assess serious game feasibility targeting autistic children?

Rational: The primary focus of this research question is to design and develop an architecture to integrate those logs (game logs and emotion logs) in order to assist game developers evaluating their serious games.

RQ3 To what extent Game Verification Tool GVT helps developers to improve their games targeting autistic children?

Rational: GVT should be tested by developers to assess the functionality, usability, and the benefits of using GVT to analyse serious games targeting autistic children. Hypothesis: Using the GVT helps serious game developers to identify game's element that triggers high stress level for children with ASD.

The research questions were formulated by implementing the method proposed by Wieringa 2014. RQ1 is considered knowledge question, while RQ2 and RQ3 are technical and empirical question, respectively.

1.3 Research Methodology

The inherent character of this research lends itself to employing the design science methodology (ibid.), manifested through the creation of a novel artifact known as the Game Verification Tool (GVT). It consists of as follows (can be seen in Figure 2):

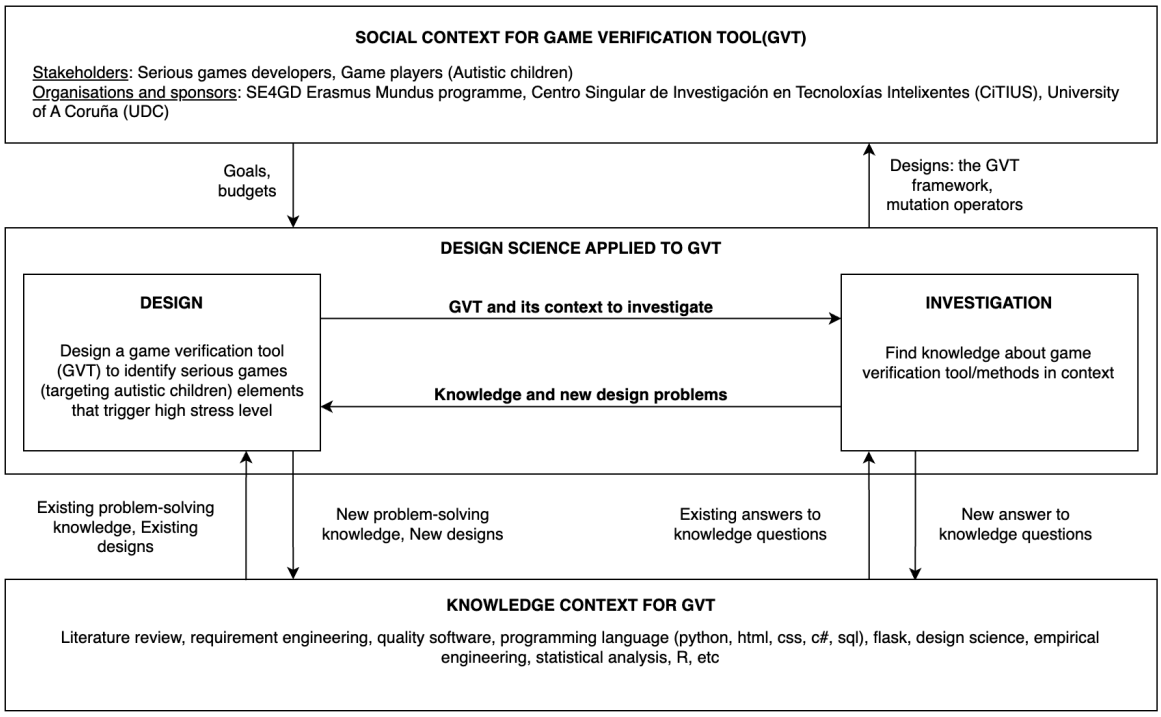


Figure 2: A framework for design science of GVT, adapted from Wieringa 2014

Additionally, the goal structure of a design science research of GVT can be seen in Figure 3.

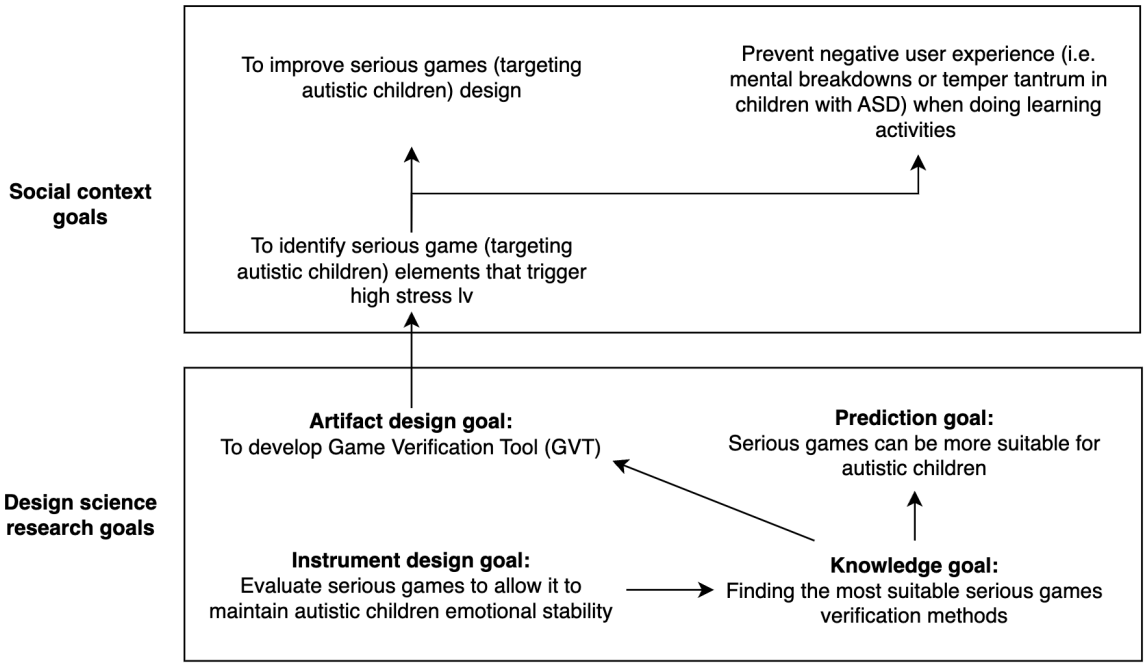


Figure 3: Goal structure of a design science research project

It was found that it is plausible to integrate game player logs and emotion logs to develop serious games verification tool (Nacke 2015; Anolli et al. 2010). However, it is found that some serious games targeting autistic children do not generate player logs (Maria Kellidou et al. 2021a; HA 2012). In addition, considering the limitation of autistic children in expressing their emotion, there is a need to find a suitable solution in which does not disrupt their emotional stability.

For these reasons, the present study focuses on the following concerns:

- Finding the suitable device which able to detect human emotion (in this case stress level) in a way that unintrusive to be applied for children with ASD.
- Proposing a standard for emotion logs structure.
- Proposing a standard for game player logs structure.
- Developing applicable serious games verification tool which fits those requirements.

The result gained from the first three concerns can be used as a key factors for identifying the solution for the fourth concern. Single-case mechanism experiments method was completed after the fourth concern was solved.

1.4 Sustainable Development Goals (SDG) of GVT

Assuming that GVT is used by majority of serious games developers whose autistic children is their target, GVT will be able to address these following United Nation SDG goals (n.d.) in social aspect:

- Quality education (SDG no 4): GVT helps developers to improve their serious games to be suitable in maintaining autistic childrens's emotional stability when doing learning activities by playing the serious games, therefore, it participates to enhance the quality of education for children with ASD.
- Reduce inequalities (SDG no 10): by providing the more suitable learning activities options for children with ASD, GVT participates in reducing inequalities and provides inclusiveness.

In addition, the sustainability awareness framework related to the aspects: social, individual, environmental, economic, and technical of the GVT are presented into Sustainability Awareness Diagram (SusAD) (Duboc et al. 2020) as can be seen in Figure 4. The black arrows represent the benefits while the red arrows represent the downsides. The colors of the cards only represent the aspect they (the cards) are in and has no further meaning.

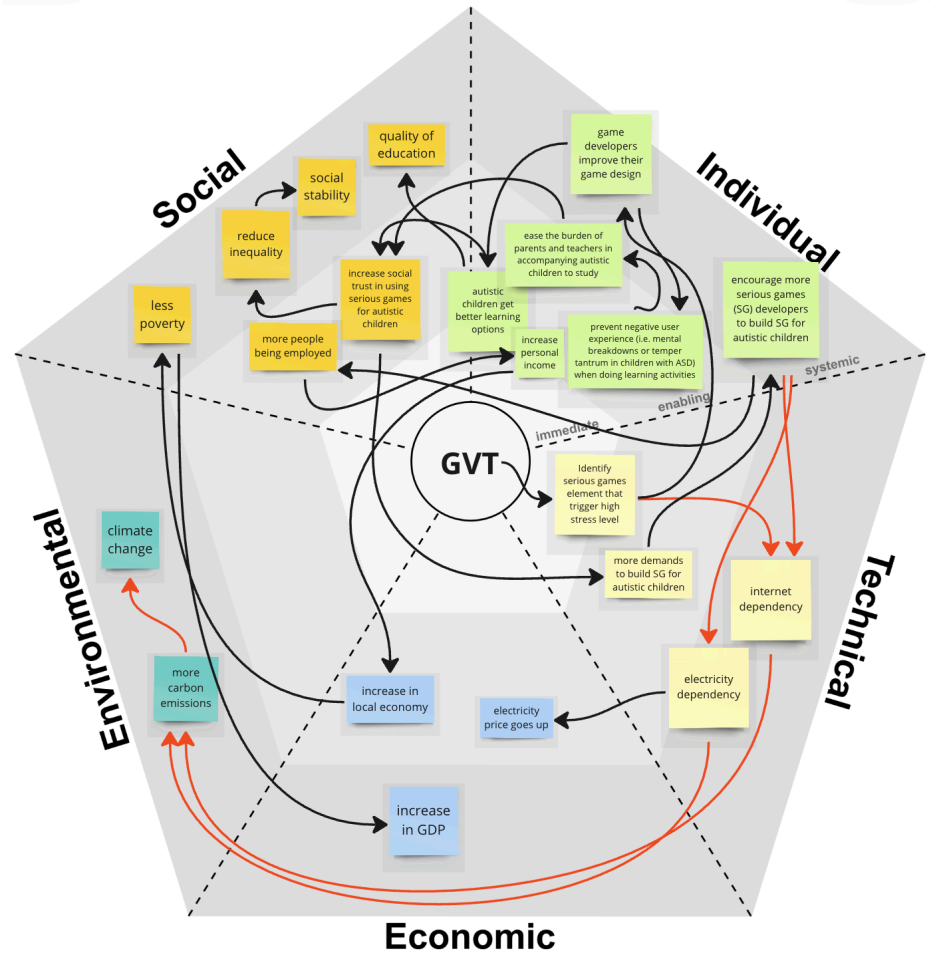


Figure 4: SusAD of the GVT

1.5 Structure of the Thesis

The structure of this work is as follows: In Section 2, the relevant findings are discussed, and the process of searching for and analyzing those findings is presented. Additionally, a selection of related findings is analyzed within the section. Section 3 presents solution design of GVT including its architecture and the data analysis process within it. Section 4 delves into the implementation. While GVT testing and evaluation process are discussed in Section 5. Lastly, Section 6 concludes the work, addressing its limitations and providing recommendations.

2 RELATED WORK

Numerous methods for Verification and Validation (V&V) have been suggested to aid game developers in creating more inclusive games. In this section, a selection of previous methods or technology used to assess and verify game components are highlighted. This section aims to address RQ1 along with its sub-research questions.

2.1 Literature Review Process and Analysis

First of all, there is a need to know the differences between Game and Serious Game (SG). According to Suits 1967, the goal in playing game is to win with rules to follow. While according to Connolly et al. 2012, playing games is for entertaining and as leisure activities. In "Serious games: Mechanisms and effects" book (Ritterfeld et al. 2009), the purposes of playing SG is to learn, to train, and for behaviour change. Means that the main goal of SG is for educational purposes.

In this literature review, previous technologies or methods that are used for games verification tools are collected along with the verification methods for serious games. The main purpose is to find applicable methods for verifying serious games targeting autistic children. Figure 5 shows the intersection between games and serious games along with the possibility for using the same verification techniques.

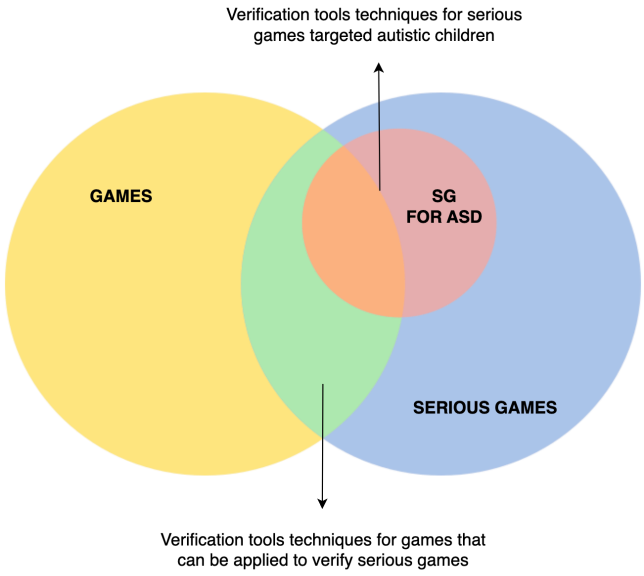


Figure 5: Existing games verification methods intersection

In the process of collecting literature, the keywords and search query used are shown in Table 1. Table 2 shows the database and specific search query used for searching the sources. There are three main databases used to gather literature: IEEE, ACM, and Springer. Additionally, snowballing method (Wohlin 2014) was also used to gather more resources. Study selection process is shown in Table 3. Using the selection process, the literature can be narrowed down before being analyzed. Furthermore, there are three main questions used to assess the findings. Those three questions are shown in Table 4, along with the search topic to filter related works. Meanwhile, Table 5 shows the score for assessing the related works. The assessment score will show whether the findings include the information related to the questions in an explicit way or not. Ergo, using Table 4 and 5, all the filtered findings can be estimated, clustered and mapped into a scheme based on the cost and end-user involvement. Cost in this case means cost to conduct the evaluation process using the particular method. The cost may be related to time duration, number of participants, and cost of equipment or tools needed, while end-user may be related to expert or game player.

Table 1: Keywords and search queries.

Keywords:	Game Verification Tool, Game Verification and Validation, Serious Game Verification Tool, Serious Game Verification and Validation, Verification Tool of Serious Game for Autistic Children, Verification and Validation Tool of Serious Game for Autistic Children
General search query:	(“Verification Tool” OR “Verification Tools”) AND (“Verification” OR “Validation”) AND (“Method” OR “Technique”) AND (“Game” OR “Serious Game”) AND (“Autistic” OR “Autism”)

Table 2: Database and specific search query.

Database	Specific search query
IEEE	(“All Metadata” : verification tool) AND (“All Metadata” : serious games) AND (“All Metadata” : autistic children)
ACM	[Title: game verification tool] AND [Publication Date: 01/01/2015 TO 01/05/2023]
Springer	game+verification+tool

Table 3: Study selection.

Include	Exclude
Article to be found, accessed	Article not available
Article in English	Article not in English
Full article available	Only abstract or poster paper available
Article describes verification tool for games verification tools for serious games AND verification tools of serious games for autistic children	

Table 4: Topic and main questions to filter related works.

Topic :	Verification and validation (V&V) methods for serious games targeting autistic children)
No	Question
Q1	what method is used?
Q2	how much the cost?
Q3	who are the end-users?

Table 5: Assessment score of related works.

Symbol	Assessment Score
Y (yes)	1
P (partially)	0.5
N (no)	0

Furthermore, by implementing snowballing method (Wohlin 2014), there is a need to select set of papers for starting point. 15 papers were gathered and grouped according to categories (Game, SG, SG for ASD) and types (Guideline or V&V) for starting point of snowballing process. Table 6 presents 15 papers related to the goals. By using those findings, snowballing process was proceed. In result, 21 V&V methods for evaluating games, SG, or SG for ASD people were found and listed, in which they can be seen in Table 7 and 8. After the existing methods were listed, those of that were summarized and estimated accordingly as can be seen in Table 9. The elaboration of the methods along with the clustering scheme will be discussed in the next section (Relevant Findings).

Table 6: Starting papers.

No	Author(s) and Year	Title	Category	Type	URL
1.	Urbanek and Güldenpfennig 2019	Celebrating 20 years of computer-based audio gaming	Games	Guideline	ACM
2.	Marciano et al. 2015	Evaluating multiple aspects of educational computer games: literature review and case study	SG	V&V	ACM
3.	Tsikinas and Xinogalos 2019	Design guidelines for serious games targeted to people with autism	SG ASD	Guideline	Springer
4.	Gao et al. 2023	Guidelines of Serious Game Design for Promoting Reframing	SG	Guideline	Sagepub
5.	Carrión-Toro et al. 2020	iPlus a user-centered methodology for serious games design	SG	Guideline	MDPI
6.	Holloway and Kurniawan 2010	Human-centered design method for serious games: a bridge across disciplines	SG	Guideline	UCSC
7.	Longstreet and Cooper 2012	Developing a meta-model for serious games in higher education	SG	Guideline	IEEE
8.	Pagulayan et al. 2007	User-centered design in games	Games	Guideline	Academia
9.	Wijnand IJsselsteijn et al. 2007	Characterising and measuring user experiences in digital games	Games	V&V	tue.nl
10.	Szilas 2022	Serious Game Design in Practice: Lessons Learned from a Corpus of Games Developed in an Academic Context	SG	Guideline	ACM
11.	Alarcon-Licona et al. 2018	From autism educators to game designers: integrating teaching strategies into game design for autism education support	SG ASD	Guideline	ACM
12.	Maria Kellidou et al. 2021b	A Review of Digital Games for Children with Autism Spectrum Disorder	SG ASD	Guideline	ACM
13.	Gris and Bengtson 2021	Assessment measures in game-based learning research: A systematic review	SG	V&V	serious games society
14.	Calderón and Ruiz 2015	A systematic literature review on serious games evaluation: An application to software project management	SG	V&V	Elsevier
15.	Petri and Wangenheim 2016	How to evaluate educational games: a systematic	SG	V&V	semantic scholar

Table 7: How the related work being collected and assessed.

No	Ref	Title	Objective	Q1	Q2	Q3	Score
1.	Razali et al. 2022	Identifying and validating game design elements in serious game guideline for climate change	Identify and validate game design elements in a serious game for climate change	Y	Y	Y	3
2.	W.A. IJsselsteijn et al. 2013	The Game Experience Questionnaire	Develop a questionnaire to measure game experience	Y	N	Y	2
3.	Ryan et al. 2006	The motivational pull of video games: A self-determination theory approach	Investigate the motivational pull of video games	Y	Y	Y	3
4.	Miura et al. 2020	GAME: Game As a Measurement Environment: Scheme to Evaluate Interfaces and Game Contents Based on Test Theories	Develop a measurement environment to evaluate interfaces and game contents	Y	Y	Y	3
5.	Drachen et al. 2013	Game Analytics – The Basics	Provide an overview of game analytics and its fundamentals	Y	P	Y	2.5
6.	Hasiah Mohamed Omar et al. 2011	Methodology to evaluate interface of educational computer game	Develop a methodology to evaluate the interface of educational computer games	Y	N	Y	2
7.	Desurvire et al. 2004	Using heuristics to evaluate the playability of games	Evaluate the playability of games	Y	P	Y	2.5
8.	Nacke 2015	Games user research and physiological game evaluation	Explore physiological game evaluation in Games User Research	Y	Y	Y	3
9.	D. Johnson et al. 2018	Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ)	Validate the Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ)	Y	Y	Y	3
10.	H. Omar and Jaafar 2008	Playability Heuristics Evaluation (PHE) approach for Malaysian educational games	Develop and apply the Playability Heuristics Evaluation (PHE) approach for Malaysian educational games	Y	P	Y	2.5
11.	Pandeliev and Baecker 2010	A Framework for the Online Evaluation of Serious Games	Develop a framework for the online evaluation of serious games	Y	N	Y	2
12.	De Kort et al. 2007	Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ)	Develop the Social Presence in Gaming Questionnaire (SPGQ) to assess social presence in digital games	Y	Y	Y	3

Table 8: How the related work being collected and assessed – continued from previous page.

No	Ref	Title	Objective	Q1	Q2	Q3	Score
13.	Sweetser and Wyeth 2005	GameFlow: a model for evaluating player enjoyment in games	Develop the GameFlow model for evaluating player enjoyment in games	Y	P	Y	2.5
14.	Mohamed and Azizah Jaafar 2010	Development and potential analysis of heuristic evaluation for educational computer game (PHEG)	Develop and analyze the potential of the Playability Heuristics Evaluation for Educational Computer Game (PHEG)	Y	P	Y	2.5
15.	Mohamed, Yusoff, et al. 2012	Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG)	Perform quantitative analysis in a heuristic evaluation for the usability of educational computer games (UsaECG)	Y	P	Y	2.5
16.	Chang and T. Johnson 2021	Integrating heuristics and think-aloud approach to evaluate the usability of game-based learning material	Integrate heuristics and the think-aloud approach to evaluate the usability of game-based learning material	Y	Y	Y	3
17.	Fu et al. 2009	EGameFlow: A scale to measure learners' enjoyment of e-learning games	Develop the EGameFlow scale to measure learners' enjoyment of e-learning games	Y	Y	Y	3
18.	Bakhuys Roozeboom et al. 2017	The effectiveness of three serious games measuring generic learning features	Assess the effectiveness of three serious games measuring generic learning features	Y	Y	Y	3
19.	Denisova et al. 2020	Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS)	Develop and validate the Challenge Originating from Recent Gameplay Interaction Scale (CORGIS) for measuring perceived challenge in digital games	Y	Y	Y	3
20.	Bowman et al. 2018	Development of the video game demand scale	Develop the Video Game Demand Scale	Y	Y	Y	3
21.	Petri, Wangenheim, and Borgatto 2019	MEEGA+, Systematic Model to Evaluate Educational Games	Develop MEEGA+ model to evaluate educational games in terms of usability and player experience	Y	P	Y	2.5

Table 9: Existing evaluation methods -summary.

No	Method	Category	Cost	End-user
1.	Game Design Checklist Evaluation + Expert Review	Serious Game Design Elements	Low/Moderate	Expert
2.	Game Experience Questionnaire (GEQ)	Game Experience	Low/Moderate	Game players
3.	Player Experience of Need Satisfaction (PENS)	Motivation in Video Games	Low/Moderate	Game players
4.	Logs based on test theories with GAME (Game As a Measurement Environment)	Interface and Game Content	Moderate/High	Game players
5.	Telemetry + Users	Game Analytics	High	Expert + Game players
6.	AHP_HeGES tool	Educational Game Interface	Low/Moderate	Expert + Game players
7.	Heuristic Evaluation for Playability (HEP) + User Studies	Game Playability and Usability	Low/moderate	Expert + Game players
8.	Games User Research (GUR)/Physiological Game Evaluation	User Experience Optimization	High	Expert + Game players
9.	Game Experience Validation (PENS & GEQ)	Questionnaire Validation	Low/Moderate	Game players
10.	PHE Approach (Heuristic Evaluation)	Educational Game Playability and Usability	Low/Moderate	Expert
11.	Online Evaluation Framework	Game Effectiveness	Moderate/High	Game players
12.	Social Presence in Gaming Questionnaire (SPGQ)	Game Experience Measurement	Low/moderate	Game players
13.	GameFlow	Players Enjoyment Evaluation	Low/Moderate	Expert + Game players
14.	Playability Heuristics Evaluation for Educational Computer Game (PHEG)	Game Usability	Low/Moderate	Expert
15.	Usability of Educational Computer Games (UsaECG)	Educational Game Usability	Low/moderate	Expert
16.	Heuristic Evaluation + Think-Aloud Method	Game Usability	Low/Moderate	Expert + Game players
17.	EGameFlow	Players Enjoyment Evaluation	Low/moderate	Expert + Game players
18.	Evaluation Study, Self-report on Competences + Post-test Questionnaire	Serious Game Effectiveness	Moderate/High	Game players
19.	Challenge Originating from Recent Gameplay Interaction Scale (CORGIS)	Game Challenge Measurement	Low/Moderate	Game players
20.	Video Game Demand Scale (VGDS)	Measuring demand in video games	Low/Moderate	Game players
21.	MEEGA+	Usability and User Experience	Low/Moderate	Game players

2.2 Relevant Findings

Initial approach was to gather and group the various techniques employed for game verification into a systematic framework to comprehensively grasp their distinctions. The summary of the previous evaluation methods has been discussed in the previous section. Based on that, all those methods were mapped to compare each of them. The scheme of the clustered methods can be seen in Figure 6.

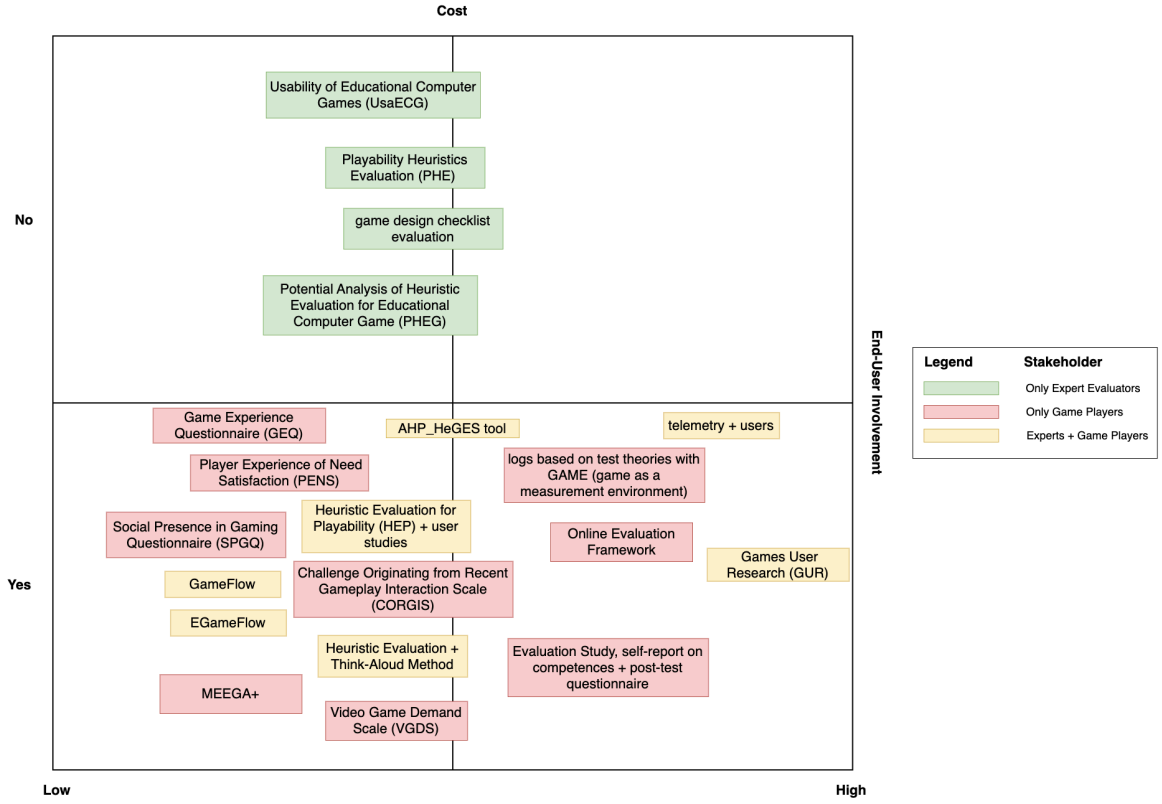


Figure 6: Mapped evaluation techniques.

Game design elements can be identified and validated according to what the game is intended for. For instance, Razali et al. 2022 conducted a research focus on validating game elements of serious games in context of climate change. In this validation method, game player involvement is not necessary, instead, experts in the field of climate change are required to validate the game design using a *game criteria checklist*. It is stated in the paper that the experts involved were given two weeks in completing the evaluation process. Thus, this approach is considered in moderate range due to the duration and experts involvement.

A method called *Game As a Measurement Environment (GAME)* was proposed by Miura et al. 2020. In the research, this method was developed to evaluate both game content and interface

based on test theories (i.e., classical test theory and item response theory). This method itself aims for effectively evaluating inclusive games. It is stated that in the evaluation process carried out using this method, high number of participants were involved along with the long duration of time. It gathered a total of 388 participants and took more than two months to do the evaluation. Considering that, GAME was mapped as relatively costly in relation of its cost.

One technique raised by Drachen et al. 2013 uses *telemetry and user* analytics technique which focuses in two aspects of game analytics: game telemetry and game metrics. By using this technique, a distance data which is valuable for game development or game research can be collected and transformed into metrics to which then being analyzed. The data processing and analytic part are considered complicated as there are eight total steps to be done. No clear statement regarding it costs. However, considering the complexity of the telemetry process, it was assumed that the cost is in high range.

Aiming to tackle in-person evaluation process, *Online Evaluation Framework* method was developed. This technique was proposed by Pandeliev and Baecker 2010, in which is designed for the online evaluation of serious games. The method particularly aims to assess game effectiveness and impact on mental fitness. The framework works as a portal to conduct a pilot study in comparing several commercially available mental fitness games with a few well-established standard board games and an active control. This method requires the involvement of game players. There is no statement related to its cost, however, evaluating the game effectiveness is challenging. To evaluate the game effect, it is required a significant length of time as well as a minimal amount of participation who wants to commit to the process (ibid.). Therefore, this method was mapped in moderate/high range related to its cost.

Methods which implement *Heuristic Evaluation (HE)* involve experts in the game evaluation process. The experts involve in the process can evaluate the game interface and design based on a set of established heuristics. Using heuristic evaluation is considered low/moderate cost since it does not require any significant financial investment. However, the cost may vary depending on the game complexity, the number of experts/evaluators involved and their level of expertise. Overall, by using this method, the evaluation can be done relatively quick which in a few hours or days and does not require extensive resources. Based on the findings, six methods are utilizing *Heuristic* evaluation: *Usability of Educational Computer Games (UsaECG)*, *Playability Heuristic Evaluation (PHE)*, *Playability Heuristic for Educational Game (PHEG)*, *Heuristic Evaluation for Playability (HEP) + User Studies*, *Analytic Hierarchy Process Based Holistic Online Evaluation System for Educational Computer Game (AHP_HeGES)*, and *HE + Think Aloud Method*.

HEP is a method proposed by Desurvire et al. 2004. This method was claimed to be suitable for evaluating general issues of the game with a prototype or mock-up in the early development phases. This method requires several evaluators to inspect the game based on player logs. The involvement of the game players is needed to fill satisfaction questionnaire as well as to play the game (to generate player logs). To gather evaluators in one same place and dispatch questionnaire to many participants may be a challenge. Thus, HEP was mapped in low/moderate in relation to its cost.

Following that, H. Omar and Jaafar 2008 proposed a method called *PHE*. This method was intended to evaluate the playability of Malaysian educational games which as well provides insights to improve game design and overall user experience. Even though implementing HE is considered easy, cheap, and fast (ibid.), there is no clear statement according to the cost in implementing this method. Hence, PHE was mapped in low/moderate range in relation to its cost, considering time duration and experts needed for evaluation process.

Extending the use of HE, Mohamed and Azizah Jaafar 2010 proposed another method with heuristic called *PHEG*. This method was intended to evaluate the usability of educational computer games. The expert evaluators involve in this method are required to predefined heuristics to assess the playability and usability of the educational games. In the evaluation process, it is only stated that there were five evaluators involved with no exact time duration. Hence, PHEG was mapped in low/moderate range in relation to its cost.

In 2011, Hasiah Mohamed Omar et al. developed a tool in which also implementing HE and extending the use of PHEG. The tool is called *AHP_HeGES* (Hasiah Mohamed Omar et al. 2011). This tool intends to work as a methodology to evaluate the educational computer games interface. This method may help in creating more effective and user-friendly educational gaming experiences. Hasiah Mohamed Omar et al. merged two evaluation techniques: PHEG and Playability Assessment for Educational Computer Game (PAEG). Addressing the limitation faced by HEP (to gather evaluators in one place), AHP_HeGES was developed as an online tool. The evaluation process of AHP_HeGES focuses on formative evaluation. Since there is no clear statement related to the cost for implementing this technique, AHP_HeGES was mapped in the moderate range since it was assumed that the tool requires the maintenance cost and the database system to save the PHEG module from expert evaluators.

Further, a method called *UsaECG* was proposed to evaluate educational game usability (Mohamed, Yusoff, et al. 2012). Only experts needed in the evaluation process. This method combines quantitative analysis with HE, in result, this method offers a systematic approach to assess and improve educational computer games usability. It is stated that in the evaluation process, around five experts evaluators (for each heuristic) were needed to be able

to detect more usability problems in the games. There is no clear statement related to the cost, hence UsaECG was mapped in low/moderate range, considering the number of variety experts needed in the evaluation process.

Another method which uses HE is the method proposed by Chang and T. Johnson 2021. The evaluation method combines *HE and Think-Aloud Method*. The method was intended to evaluate the usability of game-based learning material. Involving experts and game players, this method able to identify usability issues and enhance the learning experience. It is stated that in the evaluation process, four experts and forty-five university students were involved. Using this method, the evaluation process could be done in about three hours. Hence, this method is considered low/moderate in relation to its cost.

Aiming for integrating heuristics based on elements with a validated model in order to evaluate, design, and understand enjoyment in games, Sweetser and Wyeth 2005 proposed *GameFlow* model. The model is structured by flow, that consists of eight elements in which each element includes a set of criteria to achieve enjoyment in games. This method provides valuable insights into the factors that contribute to a satisfying gaming experience. The involvement of experts and game players are needed in the evaluation process. The number of experts, game players as well as duration for evaluation are not fixed. Those are according to the game that being evaluated. Thus, it was mapped as in range low/moderate cost.

Following GameFlow, there is *EGameFlow* method by Fu et al. 2009. This method is based on GameFlow framework which put more emphasize in measuring the enjoyment offered by e-learning games. Experts and game players are required in the evaluation process. It is claimed that EGameFlow serves as and easy and economical evaluation method for surveying learners who have used the educational game. However, it is also argued that it may be costly due to the number of experts and learners needed. Thus, side by side with GameFlow, EGameFlow is considered low/moderate cost.

Going through all the evaluation methods in detail, it is noticeable that the majority of methods which are involving game players are using questionnaire to gather players feedback related to the games. W.A. IJsselsteijn et al. 2013 proposed a *Game Experience Questionnaire (GEQ)*, which consists of: (i) core questionnaire (ii) social presence module and (iii) post-game module (D. Johnson et al. 2018; W.A. IJsselsteijn et al. 2013). A scoring guideline was also proposed to facilitate how to evaluate the questionnaire result. No statement related to its cost. However, it is assumed that duration to evaluate games using this method is relating to the number of participants (which when there is a small number of participant, the duration will be short). Besides, there is no expert or expensive equipment needed in the process. Thus, this method was mapped into low/moderate range related to its cost.

Ryan et al. 2006 also proposed game validation method using questionnaire called *Player Experience of Need Satisfaction (PENS)* (Ryan et al. 2006; D. Johnson et al. 2018). PENS has been created as an elaboration theory of Self-determination Theory (SDT). SDT itself discusses factors related to motivation (either to undermine or facilitate it). This method requires game players in the validation process. Since this method does not need much resource and the number of participant as well as the duration are not fixed, it is also considered low/moderate cost.

Since challenge is a key elements in games, Denisova et al. 2020 proposed a method to evaluate the level of players' perceived challenge in digital games. This method is called *Challenge Originating from Recent Gameplay Interaction Scale (CORGIS)*, a measurement tool amongst existing questionnaires which can be used to evaluate how players experience perceived challenges across a diverse selection of video games and among different types of players. Game players are must needed in the game evaluation process. CORGIS provides insights into players' perceptions of the game's difficulty and engagement. As the cost in using questionnaire for evaluating games is related to the number of participants involved, this method was mapped in low/moderate cost range.

Bowman et al. 2018 also proposed a questionnaire based method for measuring players' demand for video games. The method is called *Video Game Demand Scale (VGDS)*, which allows for understanding of players' motivations and preferences in gaming. As usual method based questionnaire, this method requires game players in evaluation process. It was estimated that it takes around 30 minute for each participant to complete the questionnaire before being analyzed. Meanwhile, there were in total 660 participants involved in the initial evaluation process. Considering duration and the total participants in the process, this method was mapped in low/moderate cost range.

More questionnaire based method is *Social Presence in Gaming Questionnaire (SPGQ)* which proposed by De Kort et al. 2007. This method aims to assess social presence in digital games and understanding its impact on player experience. Only game players required in the evaluation process. There were 191 participants joined to fill out the SPGQ questionnaire which estimated to take around 15 minutes to complete it. Questionnaire development typically involves the design and validation of the questionnaire, and while there may be some initial effort involved, it is generally a cost-effective method to gather data. Considering that, SPGQ was mapped as a low/moderate cost method.

Systematic Model to Evaluate Educational Games (MEEGA+) is also one of the questionnaire based methods. It was proposed by Petri, Wangenheim, and Borgatto 2019, which aims for evaluating educational games in terms of usability and player experience from the students' perspective in the context of computing education. MEEGA+ offers a structured and

comprehensive framework for assessing the effectiveness and educational value of educational games. Game players are a must required element in the evaluating process. There is no clear statement regarding the cost for this method. However, as usual questionnaire based method, the number of participants correlate with the time duration to analyse the result. Hence, MEEGA+ was mapped in low/moderate range of cost.

Based on the findings, there is a method which combining three methods into one. Bakhuys Roozeboom et al. 2017 proposed *Evaluation Study, self-report on competences + post-test questionnaire* method. This method was suggested to evaluate the effectiveness of three serious games designed to measure generic learning features. The findings from the evaluation provide insights into the games' efficacy in supporting learning outcomes. As it was mentioned before, evaluating game effectiveness is challenging. A long duration of process is needed along with the commit participants. The same happened using this method. It is stated that more than 70 participants involved and one of the evaluation case needed around two months to complete. Considering that, this method was mapped in high cost range.

Games User Research (GUR) technique was proposed by Nacke 2015. The technique is used to optimize the user experience in games and virtual entertainment products. Physiological evaluation is aimed to be a standard tool in GUR. Considering that the targets are for academic and industrial applications, it was stated that currently the physiological evaluation methods used for GUR require expensive equipment which is used primarily in a laboratory setting. GUR implements physiological responses evaluation which requires these following equipment: Cardiovascular Measures, Electrodermal Activity (EDA) and Galvanic Skin Response (GSR), Electromyography (EMG), and Electroencephalograph (EEG). With that point, the cost to apply GUR is the highest compared to other mapped techniques.

One valuable research related on how to assess emotion in serious games was found. According to Anolli et al. 2010, there are ways to evaluate emotional experience of SG users. The list of emotion assessment methods is as follows:

- Self-Report Measures of Emotions: This method is utilizing either questionnaire to be filled by participants or pictorial methods (e.g. Self Assessment Manikin Scale). These measurements are psychometrically validated and are considered non-intrusive, fast and cheap methods. Five emotion self-report measures which may be utilized are as follows:
    - *Mood Adjective Checklist (MACL)* (Nowlis 1965)
    - *Profile of Mood States (POMS)* (McNair et al. 1971)
    - *R-Multiple Affect Adjective Checklist (MAACL-R)* (Zuckerman et al. 1983)
    - *Expanded Form of the Positive and Negative Affect Schedule (PANAS-X)* (Watson and Clark 1994)

- *Differential Emotions Scale (DES)* (Izard et al. 1993)

- Physiological correlates of emotions: Utilizing physiological system which responsible for modulating peripheral functions, Autonomic Nervous System (ANS). ANS consists of sympathetic and parasympathetic branches which associate with activation and relaxation, respectively. The range of psycho-physiological measures available is vast, including *Skin Conductance Level (SCL)* (Öhman and Soares 1994) and cardiovascular correlates of emotion as follows:
  - *Heart Rate (HR)*
  - *Heart Beat Amplitude (HBA)*
  - *Heart Rate Variability (HRV)*
  - *Finger Pulse Amplitude (FPA)*
  - *Pulse Transmission times to the Finger (PTF)*
  - *Pulse Transmission time to the Ear (PTE)*
  - *Pre-Ejection Period (PEP)*
  - *Diastolic Blood Pressure (DBP)*
  - *Systolic Blood Pressure (SBP)*
  - *electromyography*
  - *respiration rate and amplitude*
  - *level of blood glucose*
  - *temperature*, etc
- Facial expressions: *Facial Action Coding System (FACS)* (Ekman and Friesen 1978; Ekman, Friesen, and Tomkins 1971; Izard 1984; Als et al. 1980; Ekman and Friesen 1978)
- Vocal-nonverbal features: change of *voice*. Following are the example of acoustic analysis of voice which may applicable to use:
  - *Computerized Speech Lab (CSL)* developed by KayPENTAX
  - *PRAAT* software fhttp://www.fon.hum.uva.nl/praat/
- Gestures:
  - *Observer XT 10* http://www.noldus.com/
  - *Theme 5.0* http://www.patternvision.com/

Notice that most of the previous techniques are aiming for general games or SG, they require the involvement of end-user in the V&V process. However, those approaches have not considered the limitation of children with mental disorder which makes the V&V process relatively uninclusive. Additionally, in relation to this master thesis work, the most closest research to evaluate games which may be applicable for evaluating serious games targeting autistic children is GUR. However, GUR implements physiological responses evaluation

which requires equipment in which those are considered obtrusive for mental disorder children especially children with ASD.

Furthermore, although many applicable emotion detection methods exist (Anolli et al. 2010), taken into account traits of person with ASD (Sucksmith et al. 2011), majority of those methods are not applicable for them. Self-Report measures of emotions may not be reliable if we expect children with ASD to fill the questionnaire or use pictorial methods. Let alone filling the form, it is already a difficult task for them to even express their emotion. Taken into account that particular trait along with easily feel anxious, implementing facial expression, gestures, and voice detection or analyzer, are not plausible as well. Further, utilizing physiological system measurements, may not be the best solution for this case, since almost all the physiological system measurements require a person to wear on-skin equipment, which is considered intrusive for autistic people.

Taken into account the goal of this work which focuses in analyzing game elements to maintain players stress level, it is a must to find a solution for detecting autistic players' stress level in unintrusive way. There is a need to diminish using equipment that may trigger high stress level for players when carrying the testing and evaluation process. On the other hand, a simple V&V tool may be beneficial in cost and time related aspects to assist not only for industrial application, but also small start-ups or individual game developers who has interest in developing serious games for autistic children to validate their games. Doing so, this master thesis work proposes a technique which only entails a simple wearable stress detector for players to use. Hence, the technique is reconnoitred as unobtrusive for autistic players. In summary, GVT has been developed to be a simple but beneficial low-cost serious games (targeting autistic children) verification tool.

Further, based on the literature review and research findings, it is possible to integrate game player logs and emotion detector to develop verification tool. Therefore, next chapter will discussed about the solution design in developing GVT along with the architecture of GVT.

3 SOLUTION DESIGN: THE GAME VERIFICATION TOOL

In this chapter, the analysis and design process will be presented. It is done by drafting the GVT functional and non-functional requirements and specifying minimum requirements needed for creating **emotion logs** and **serious games logs**. Moreover, a structure for both logs is also defined. Regarding the design, the GVT architecture is presented, where the data manager module and analysis process are described with detail.

3.1 GVT Requirements

Table 10 shows the composed **functional requirements** for GVT. There are three functional requirements for GVT to fulfill. GVT must be able to integrate emotion detector and game players logs. This is the baseline requirement for GVT to have. By having this requirement, ensures GVT to record emotional state (in this case stress level) of the game players in every game players logs. Following this requirement, GVT should be able to locate the potential scenes that trigger high stress level for players. In order to assist game developers in improving their SG, GVT must be able to identify logs in every scenes of the game with high stress level. Thus, it will give a chance for the developer to locate particular logs and encourage them to improve that particular elements in those scenes. Having just two aforementioned functional requirements is not enough for GVT to provide valuable information for the SG developers. Thus, GVT should be able to analyze the data received from ED and games logs by giving an assessment result as a game feasibility score. The goal is to analyze whether the SG targeting autistic children really is suitable for them or not by considering players stress level. The score of the SG feasibility generated by GVT can provide an insight information for the SG developers related to the suitability of their games for those of that targeted players.

Table 10: Functional requirements of GVT

Functional
Integrate emotion detector and game players logs
Identify potential scenes that causing stress for players
Analyze input data and give an assessment result as a feasibility score

Considering **non-functional requirements**, it is required for GVT to have usable characteristic for GVT's users which is including to have atrributes: appropriateness recognizability, learnability, and operability along with availability and maintainability. The usabil-

ity requirement is needed in order for GVT's users to effortlessly operate it. According to (Condori-Fernandez and Lago 2018), having appropriateness recognizability attribute of usability means that the users have the ability to assess the suitability of a system for their requirements prior to its implementation. Learnability means that the system can be utilized to accomplish predetermined learning objectives related to system usage. Additionally, operability attribute of usability means that the system possesses characteristics that facilitate ease of operation and control. The availability attribute indicates that the GVT should be available and accessible when it is needed to be used. Since the GVT aims to be a simple but useful verification tool, having maintainability characteristic is a must. Thus, the system can be easily maintain and modify to fulfill the needs.

3.2 Definition of Structured Logs for GVT

Since GVT is a logs-based verification tool, there is a need for GVT to standardize structure of the emotion logs and the SG player logs in order to make them eligible to be integrated and analyzed by GVT. This subsection provides a detailed explanation of the prerequisites for emotion and game logs. The discussion encompasses the selection of a suitable reference for establishing a baseline in the design of emotion logs structure for GVT, as well as the author's proposal of a minimal structure for serious game logs.

3.2.1 Emotion Logs

Analysis process includes choosing what emotion detector to use for the reference in designing and developing GVT. Taken into account the original purpose of GVT, it is a must to make sure that all devices which in contact with the autistic players are not considered hostile. Thus, sensors used by ED should be a simple and comfortable wearable device. In collaboration with Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), emotion logs preconditions was designed based on an unobstructive device from a research done by Suni Lopez et al. 2019. According to their research work, the **stress detector** has accuracy of 79.17%, a precision of 60% and a recall of 50% . Thus, the device is considered reliable for stress detector and accurate for stress detection. Using those information, the proposed minimum requirements for the emotion detector to ensure a connection with GVT are presented in the Table 11.

Table 11: Functional requirements of ED

Functional
Detect emotional level of user (particularly stress level) from wearable sensors
Analyze the value received from sensors by giving range of value of stress level
Able to send the stress level value to web server application by inputting endpoint address

Based on that, GVT was designed to accept any ED input with the minimum structure (in JSON format) as seen in Figure 7. The ED by Suni-Lopez et al. is using 5 points scale of stress level measurement, where 1 represents the absence of stress (relaxed), and the present of stress begins at point 4, with the highest level corresponding to 5.

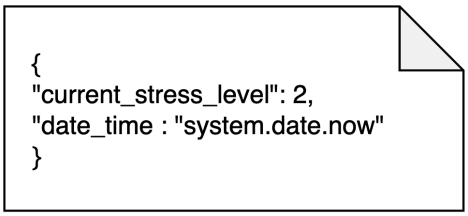


Figure 7: Minimum structure of emotion logs in JSON format

3.2.2 Serious Games Logs

Author proposes that serious games targeting autistic children should generate game player logs which preferably in JSON format in order to be easily verified and validated by serious games verification tool (GVT). Author recommended that the minimum structure for the logs is having attributes as follows:

1. **Time**: yyyy-MM-dd HH:mm:ss
2. **Scene**: scene_name
3. **Asset**: asset_name
4. **Events**: click / drag / drop / record / etc
5. **Task**: success / failed / right / wrong/ etc

The proposed minimum structure was defined by analoging school subject materials' problems as a game. To describe it in details, let us use Mathematics as an example. Mathematics has many branches, for instance: Arithmetic, Algebra, Geometry, Trigonometry etc. These branches can be seen as **Scene** in serious games to indicate which part of the game the players are studying/playing. To understand the subject better, students are expected to solve problems related to that subject materials. **Asset**, can be seen as everything related with the problems. Asset can be questions or answer choices. Meanwhile, there are many ways to solve problems. Some problems may provide multiple-choice answer to select and some others may require students to write down the answer in details (in mathematics, they may require to write the correct formula and going through the calculating process). **Events**, reflects the way the problems are solved. Events can be mouse manipulation, or how the players are expected to do to solve the problems. Additionally, assessment is needed to know whether the answer is correct or incorrect. **Task** (in which the simplify of task completion), represents

those assessment. Task is giving the evaluation on how well the players complete the tasks given. Further, there must be an expected time duration to finish math exam, likewise, in this case, **Time** indicates the time when students/players start to work on the exact part of those of that learning activity, in which later will be calculated to get the time duration.

In summary based on that analogies, game logs time variable should represent the time for every triggered events. Scene should indicate chapters of the game which are composed of assets, tasks, and events. Meanwhile, asset is property that construct the game (for example: button, image, etc). Task represents how well the players complete the game challenge, while events are how the player finish the task using external input (for example mouse manipulation). The proposed meta model and the example of JSON format file of the logs can be seen in the Figure 8 and 9 respectively.

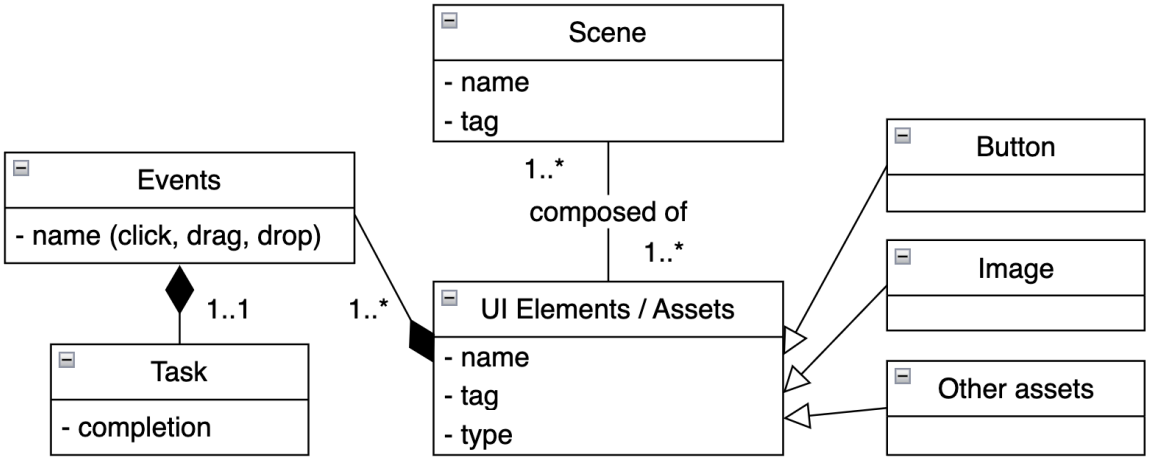


Figure 8: Meta model of the proposed serious games logs

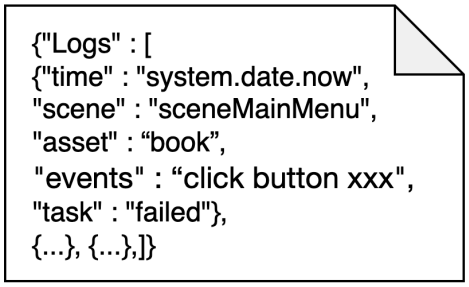


Figure 9: Minimum structure of proposed game logs in JSON format

3.3 GVT Architecture

Prior to the GVT development process, GVT architecture must be constructed. The GVT architecture employs a client-server model, which is depicted in Figure 10. This diagram illustrates the process by which a GVT server distributes resources for analyzing data collected from a network, where both the game to be verified and the GVT desktop are hosted.

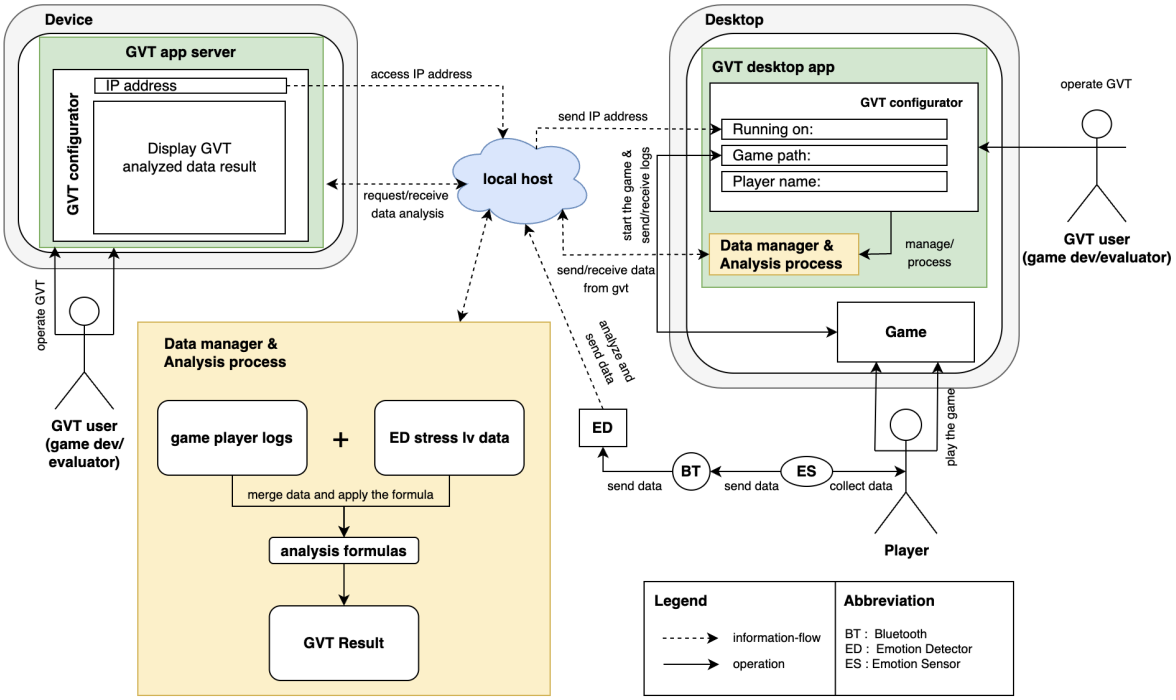


Figure 10: GVT architecture

The **GVT desktop application** consists of **GVT configurator** module, which offers the following functionalities: (i) sets the local host IP address, (ii) gets the game path, and (iii) gets the game's player name. Modules (ii) and (iii) require input from **GVT user** to set the game path and input the player's name, respectively, for initializing the GVT. Upon initialization of the GVT desktop app, the chosen game within the specified path is launched automatically for verification. The **Data manager** component starts monitoring the player logs, while the player's emotion logs (stress level values) are collected by the **ED** device via **BT** which then is sent to the server end-point during gameplay. Subsequently, the collected data undergoes an **Analysis process** for further examination. Additionally, the **GVT app server** can be operated from another **Device** by the GVT user. Operating the GVT app server is necessary to display the results obtained from the Data manager and the Analysis process in a suitable manner.

3.3.1 Data Manager and Analysis Process

Data manager and analysis process is one of the key modules of the GVT. The main purpose of this module is to manage and analyse data into and from database. From information received by GVT desktop app which is configured by GVT user, the information then being processed by data manager and analysis process module to which the data being managed and analysed. Managed and analysed data is then being shown to GVT web app.

Data Manager

Data manager's main functionalities are related with the logs integration and GVT database modification. Information received by GVT from GVT desktop app and ED, are sent to the database accordingly. In result, through the data manager, GVT database can record and save the emotion logs and game logs in the most suitable way to be analysed later on. Relating to the data modification, information received by GVT from GVT web app, provokes data manager to do database modification. In result, the data can be analysed accordingly.

Giving the example in what the data manager does in the integration process, raw data from game logs and emotion logs can not be sent directly to the GVT database without any initial configuration. There should be a "key" that allows them to be integrated inside the database. GVT uses **time** and **player name** as the key, allows it to integrate the logs from ED and the serious game into one records. The same process happens when the data manager does the database modification. Start from receiving the information from user input via GVT web app, data manager filters the database accordingly, allowing the analysis process happens afterwards.

Analysis Process

In order to provide game developers with valuable data analysis, GVT necessitates the examination of data generated by player logs (both from the ED and game) using the formula outlined in Table 12. By analyzing player logs and emotion logs, GVT can yield new data such as **time spent**, **time accumulation**, **total time in scene**, and **average stress in scene**. Additionally, for multi-records analysis, by comparing data from multiple players logs, GVT can obtain a feasibility result of the analysed games. By extending the use of the existing formulas, GVT can calculate the multiple players logs and produce the assessment result as **overall stress level** for the game, along with **overall time game play**. These results are gained by initially calculated the avg stress level each scene and avg total time each scene of multiple players logs. For calculating the multi-players records, GVT will grouping all the records from those multiple players into the same scene. After all the scenes are grouped, it calculates the **avg stress level each scene** and the **avg total time** in those each scene using the same logic used to analyse single-player records. Note that the more the games being

analysed (by testing it with multiple players), the more players logs GVT obtains. In result, the more reliable the game feasibility score will be.

Table 12: Formula used in GVT

Value	Formula
Time Spent	ts E0 = *(tE1 - tE0)*
Time Accumulation	ta En = $\Sigma((tE1 - tE0) + (tE2 - tE1) + (\ldots) + (tEn - (tEn - 1))$
Total Time in Scene	tScene = $\Sigma(tsE0 + tsE1 + .. + tsEn)$
Avg Stress in Scene	$\Sigma(((sE0s * tsE0s) + (sE1s * tsE1s) + (sE2s * tsE2s) + \ldots + (sEns * tsEns)))/\Sigma(tsE1s + tsE2s + tsEns))$
Min Stress Lv in Scene	min = (min(stress_lv))scene
Max Stress Lv in Scene	max = (max(stress_lv))scene
Mode Stress Lv in Scene	mode = (mode(stress_lv))scene
Quartiles Stress Lv in Scene	quartiles = (quartiles(stress_lv))scene
Standard Deviation Stress Lv Scene	std = (std(stress_lv))scene

The duration between the current and next event in the game is referred to as **time spent**. This is calculated by subtracting the current event time from the next event time. **Time accumulation** represents the total duration from the first event to the current event. It is calculated by subtracting the previous event time from the current event time and adding the result to the accumulated time. **Total time in scene** denotes the overall duration spent in a specific scene. This is calculated by summing the time spent on each activity within that scene. On the other hand, **average stress in scene** represents the average stress level in a particular scene. This is calculated by multiplying the stress level by the time spent in each activity within the scene and dividing it by the total time in the scene. Meanwhile **Min**, **Max**, **Mode**, **Quartile**, and **Standard Deviation Stress Lv in Scene** represent the minimum, maximum, mode, quartiles, and standard deviation of stress level values in particular scenes, respectively. Those statistics calculation[1] are calculated using statistical function in accordance to all the stress level of every events in each scene .

There are some reasons taken into consideration in proposing those aforementioned formulas. According to Allman (Casassus et al. 2019), the behaviour traits of people with ASD such as lack in social function, adaptive cognitive and behaviour highly affected by the deficiency

[1]The inclusion of these statistics were included during the second iteration of the GVT tool development (after usability evaluation)

of sensitivity to perceive time, including the events duration and time between them. Additionally, since the GVT was developed to give feasibility score of SG that targeting autistic children, there was an assumption that the game developers do specifically target those of that children with the similar characteristics and behaviour.

By applying this formula in conjunction with game logs, developers can identify the specific scenes that induce a high average stress level based on time-related aspects. These scenes consist of various elements such as assets, events, and tasks. Thus, by only analysing the scene, it is considered sufficient to provide insights into areas that require improvement to make the scene less stressful.

3.4 UI Designs

Prior to the implementation process, GVT UI was designed. Two main designs of GVT includes: GVT desktop app and GVT web app. In this section, the designs are elaborated with their main functionalities. The designs of GVT desktop app and GVT web app will be discussed in separate subsections.

3.4.1 GVT Desktop App

GVT desktop app was designed to run the game as well as the web app and the database directly when the GVT is started. Figure 11 shows the UI design for GVT desktop app.

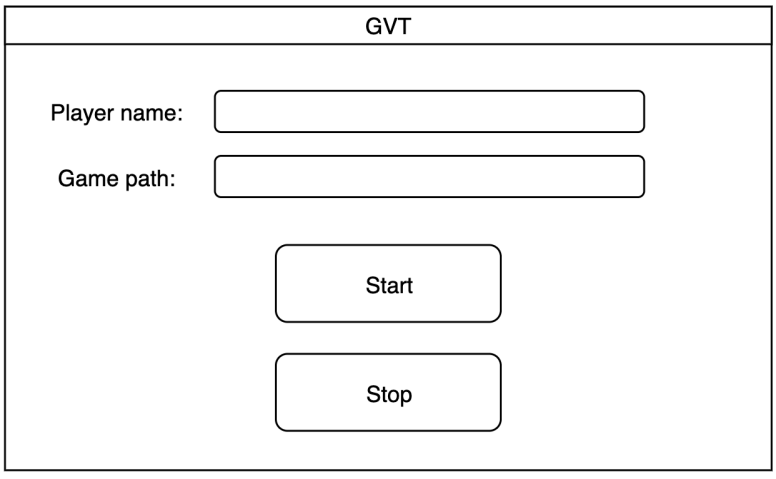


Figure 11: UI design of GVT desktop app

Going into details, in the mock-up, there are two entry columns in GVT desktop app. Player's name entry is needed to differentiate players when there is multi game testing happens at the

same time. Game path entry uses for locating the game in which game players logs usually resides in the same path as the game. Start button will trigger the run of the GVT server and database as well as the game, while the stop button will stop all the GVT process.

3.4.2 GVT Web App

GVT web application has more dynamic UI designs. There are four main navigation menus: Home, Single Analysis, Multi Analysis, and About. Home menu represents the front page of the GVT web app which intends to direct the GVT users into the main analysis menus: Single Analysis and Multi Analysis. Single Analysis menu itself is intended to analyse only single data record in detail. Meanwhile, Multi Analysis page is intended to analyze multi data records in order to get final result and comparable analysis of each record. Additionally, About menu acts to give a brief information about the GVT.

Generally, **Home** menu intends to act as a front page which showing all the menus that the GVT has. Therefore, there are only three buttons which each of them link the GVT user to the other menus. There is also menu bar on top of the page to navigate and show the active page. The design of the Home page can be seen in Figure 12.

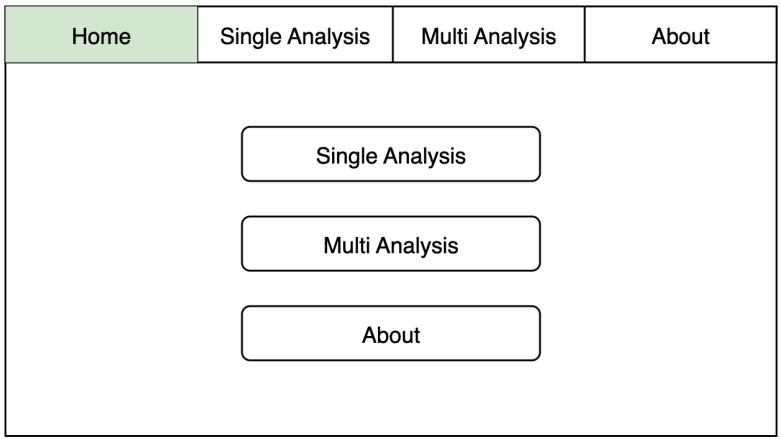


Figure 12: Home page UI design of GVT web app

The second main menu is **Single Analysis**. This page is intended to show an analysis for single-records. The design of the Single Analysis page can be seen in Figure 13. GVT users are allowed to select the name of a player that is available inside GVT database from the player name select drop down options. Following that, the date of play will show the available date according to the name chosen from the selection options.

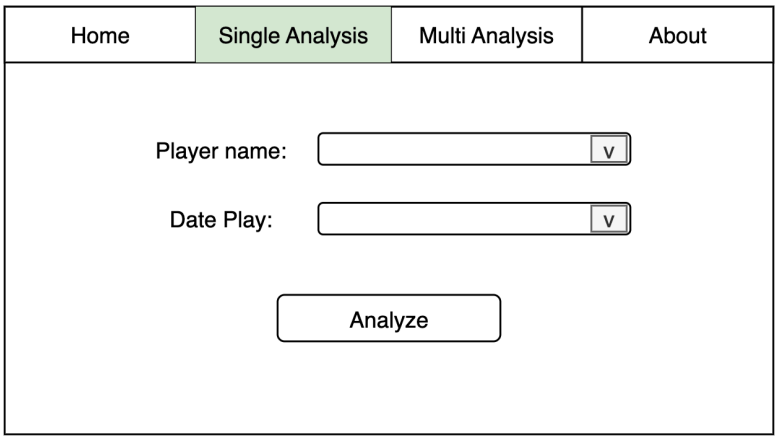


Figure 13: Single Analysis UI design of GVT web app

The third main menu is **Multi Analysis**. This page is when the multi testing players records can be analysed. As it is mentioned before that in using GVT, "the more records are analysed, the more reliable score of game feasibility is resulted". This particular page is intended to do that analysis for giving final result score of SG feasibility for autistic children.

The UI design of the Multi Analysis page is depicted in Figure 14. Implementing the same functionalities as in the Single Analysis page, users are allowed to select the available Player Name and Date Play from the database. However, to extend that function, users can add more than one player records to be analysed.

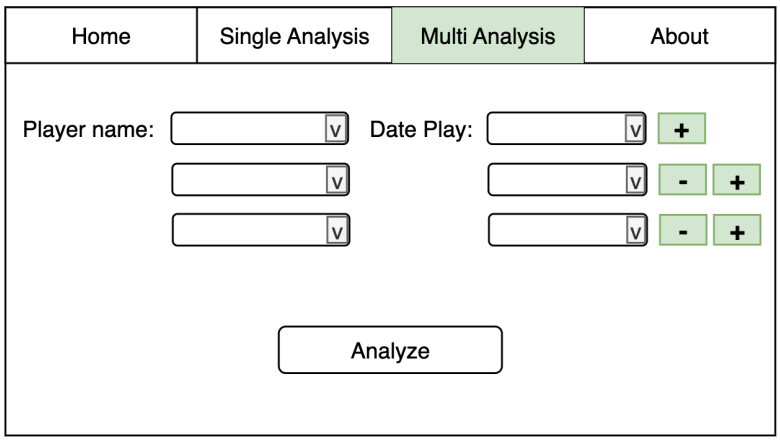


Figure 14: Multi Analysis UI design of GVT web app

Last but not least, **About** menu. This menu gives a brief information about the GVT. Who the persons are behind it and what the work is about. The UI design can be seen in Figure 15.

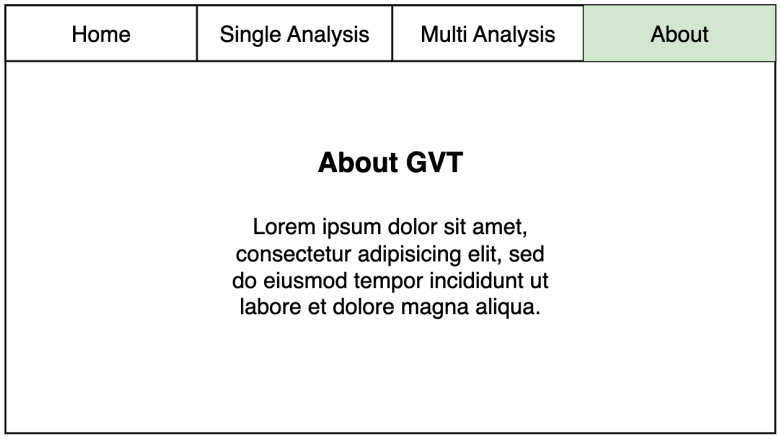


Figure 15: About page UI design of GVT web app

4 THE GVT IMPLEMENTATION

In this section, the recap on how the solution design and analysis process was implemented to build GVT is written. Selected technology for GVT is discussed. Further, proposed script for game logs is revealed along with the final look of GVT after the designs were realized.

4.1 Selected Technology

In this section, selected technology for GVT is discussed. Prior to developing GVT, it is necessary to preselect the suitable technology and define the condition needed for external data inputs for GVT, which are emotion logs and game logs. Figure 16 provides an illustrative representation of the GVT blueprint, presenting various technology options that could be employed.

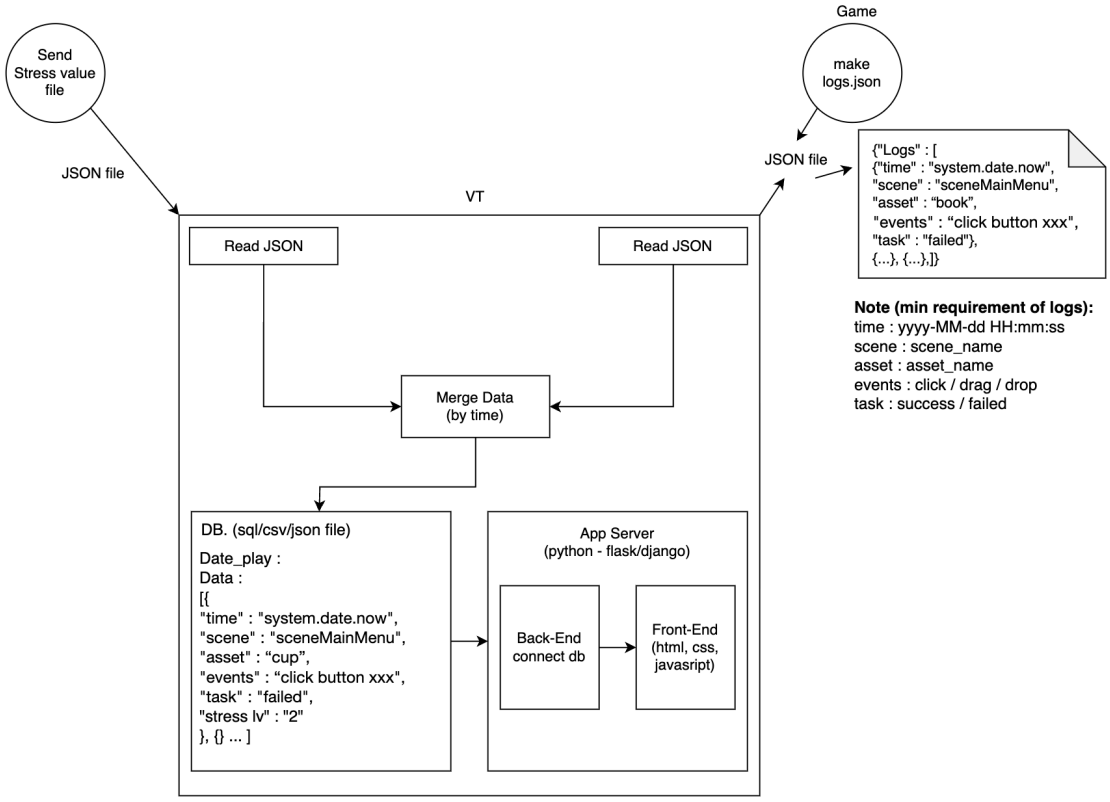


Figure 16: GVT blueprint

Considering those options and GVT requirements, the main programming language was decided. Having an abundance of open-source package libraries, along with its simplicity,

practicality, and explicitness (n.d.), making python to be the best option compared to other programming languages for developing GVT. Those of that python traits help addressing the maintainability of the GVT non-functional requirement. Python tkinter was utilized for standard desktop Graphical User Interface (GUI), while python flask was used for building app server which connects the back-end and front-end with the database. The database itself was established based on MySQL, therefore, PyMySQL was implemented. On the other hand, html and css language was used to build the front-end part in which GVT app server was deployed.

JSON was chosen for the preferably file format for GVT by considering that it is the most common format used for electronic data-interchange (n.d.), particularly in the data transmission within web applications. The goal is to make GVT to be able to handle any data format to address its usability requirement. However, to ensure a smooth data transmission, JSON is a more suitable format for GVT.

4.2 Proposed Game Logs Script

After recommending the structure for game logs, author proposed particular script that may be reusable to be implemented for other serious games built using Unity software. The fully script shown in Appendix A Listing A.1 is the code that used by author to modify Gemas for generating the logs. The purpose of the script is to enable the creation and storage of JSON logs for a game using Unity C#. It records various data and events that occur during gameplay and saves them in a JSON file called “PlayerGameLogs.json”. The elaboration of the script will be discussed in the following snippets.

Listing 4.1 presents the snippet of the script in which showing the used of necessary libraries.

```
1 using System;
2 using System.Collections;
3 using System.Collections.Generic;
4 using System.IO;
5 using UnityEngine.SceneManagement;
6 using UnityEngine;
7 using UnityEngine.EventSystems;
8 using UnityEngine.UI;
```

Listing 4.1: MakeJsonLogs necessary libraries

Listing 4.2 represents the structure of a single log entry. It contains several public fields that will be recorded for each log entry:

- **logTime**: A timestamp representing the current date and time when the log is recorded.
- **scene**: A string that stores the name of the scene in the SG.

- **asset**: A string that stores the name of the SG object asset
- **events**: A string to store the specific event that occurred (e.g., mouse manipulation).
- **task**: A string representing the task or action associated with the logged event.

```
[System.Serializable]
public class Logs
{
    public string logTime = System.DateTime.Now.ToString("yyyy-MM-dd HH:mm:ss");
    public string scene;
    public string asset;
    public string events;
    public string task;
}
```

Listing 4.2: MakeJsonLogs logs class

Listing 4.3 represents the structure of the log list. It contains a public array of *Logs* objects. The intention is to store multiple log entries in an array.

```
[System.Serializable]
public class LogsList
{
    public Logs[] logs;
}
```

Listing 4.3: MakeJsonLogs LogsList class

Listing 4.4 shows class members that create instances of *LogsList* and *Logs* objects to hold the log data.

```
public LogsList myLogsList = new LogsList();
public Logs myLogs = new Logs();
```

Listing 4.4: MakeJsonLogs class members

Listing 4.5 shows a method that serializes the *myLogsList* object into a JSON string using *JsonUtility.ToJson()*. The true argument prints the JSON. Then, it writes the JSON data into a file named “PlayerGameLogs.json” located in the Unity project’s “Asset” folder.

```
public void MakeFile()
{
    string logsJson = JsonUtility.ToJson(myLogsList, true);
    File.WriteAllText(Application.dataPath + "/PlayerGameLogs.json", logsJson);
}
```

Listing 4.5: MakeJsonLogs MakeFile() method

Listing 4.7 presents method used to add a new log entry to the *myLogsList* object. It takes two parameters: events and task, which represent the event description and the task associated with the log. In this method, the script reads the existing JSON data from the file "PlayerGameLogs.json" and deserializes it back into the *myLogsList* object using *JsonUtility.FromJson<LogsList>()*. It then creates a new log entry (myPlayerLogs) with the current timestamp, scene name, asset name, event description (events), and associated task (task). The method then adds the new log entry to the existing *myLogsList* by creating a temporary list (myLogsListTemp) to hold the old log data. It then resizes the original log array to accommodate the new entry and copies the old data back. Finally, it calls the *outputJSON()* method to write the updated *myLogsList* back to the JSON file.

```
public void AddRecord(string events, string task)
    {
        //read json from text file
        string str = File.ReadAllText(Application.dataPath + "/
    PlayerGameLogs.json");
        myLogsList = JsonUtility.FromJson<LogsList>(str);
        //collect data from input fields
        Logs myPlayerLogs = new Logs();
        myPlayerLogs.logTime = System.DateTime.Now.ToString("yyyy-MM-dd
    HH:mm:ss");
        myPlayerLogs.scene = EventSystem.current.
    currentSelectedGameObject.tag;
        myPlayerLogs.asset = EventSystem.current.
    currentSelectedGameObject.name;
        //myPlayerLogs.events = EventSystem.current.
    currentSelectedGameObject.name + " is clicked";
        myPlayerLogs.events = events;
        myPlayerLogs.task = task;
        //make temp list to hold old data
        LogsList myLogsListTemp = new LogsList();
        myLogsListTemp.logs = myLogsList.logs;
        //make original longer
        myLogsList.logs = new Logs [myLogsListTemp.logs.Length + 1];
        //copy old data across
        for (int i = 0; i < myLogsListTemp.logs.Length; i++)
        {
            myLogsList.logs[i] = myLogsListTemp.logs[i];
        }
        //add new data
        myLogsList.logs[myLogsListTemp.logs.Length] = myPlayerLogs;
        //write data
        outputJSON();}
```

Listing 4.6: MakeJsonLogs AddRecord() method

Listing 4.7 shows a method that is responsible for serializing the updated *myLogsList* object into a JSON string and writing it back to the "PlayerGameLogs.json" file.

```
public void outputJSON()
{
    string strOutput = JsonUtility.ToJson(myLogsList);
    File.WriteAllText(Application.dataPath + "/PlayerGameLogs.json",
    strOutput);
}
```

Listing 4.7: MakeJsonLogs outputJSON() method

To summarize, the script (Listing A.1) allows the creation and modification of a JSON log file with event data that has timestamps in a Unity game. The JSON file is initialized by the *MakeFile()* method, and new log entries are added to the existing JSON data by the *AddRecord()* method. The log entries have timestamps, scene names, asset names, event descriptions, and tasks.

The example on how to call MakeJsonLogs class from main menu is shown in Listing A.2. Author used Indonesian language for the class name and Buttons. The meaning of those of that are as follows: menu utama (main menu), warna (color), kenal bangun (learn shapes), ayo bersikap (to act), keluar (exit), kembali (back), bantuan (help), panel bantuan (help panel). Meanhwile, Listing A.3 presents an example for generating logs for task completion. The code is for drag and drop task which if the player drop answer in the correct place the task completion will be success and vice versa. Some of words are in Indonesian language. The meaning of those of that are as follows: soal (question), jawab (answer), benar (correct/success), salah (incorrect/fail). There are codes in those Listing (A.2 and A.3) scripts that need to be taken into account.

The code in Listing 4.8 must be called to declare and initialize a variable named *mj* of type *MakeJsonLogs*. It can be used to interact with and access the functionalities provided by the *MakeJsonLogs* class. It allows the script to utilize the logging and data serialization capabilities of the *MakeJsonLogs* class to record and manage game logs.

```
public MakeJsonLogs mj = new MakeJsonLogs();
```

Listing 4.8: Example of calling MakeJsonLogs

Additionally, the code in Listing 4.9 must be called to create or added a new log entry inside the "PlayerGameLogs.json".

```
mj.AddRecord (" Player goes to closing scene ", "");
```

Listing 4.9: Example of calling AddRecord function of MakeJsonLogs

Note that all of the scripts related to the game should call MakeJsonLogs in order for them to generate logs for each events in the scene. Author is looking forward for the code to be reusable for other serious games. The result of generated game player logs after implementing those scripts can be seen in Figure 17, which the JSON file of the game logs is generated in the same path as the game. Figure 18 shows the example when game Gemas generates game logs into the same path as the game.


```
{"logs":[{"time":"2023-05-12 20:06:08","scene":"Opening
    Scene","asset":"Button_mulai","events":"Player goes to main
    menu","task":""},{"time":"2023-05-12 20:06:12","scene":"Main
    Scene","asset":"Button_shapes","events":"Player goes to study shape
    menu","task":""},{"time":"2023-05-12 20:06:15","scene":"Shapes
    Scene","asset":"Button_persegi","events":"Player clicks
    square","task":""},{"time":"2023-05-12 20:06:17","scene":"Shapes
    Scene","asset":"Button_segitiga","events":"Player clicks
    triangle","task":""},{"time":"2023-05-12 20:06:18","scene":"Shapes
    Scene","asset":"Button_persegi panjang","events":"Player clicks rectangle
    ","task":""},{"time":"2023-05-12 20:06:19","scene":"Shapes
    Scene","asset":"Button_lingkaran","events":"Player clicks
    circle","task":""},{"time":"2023-05-12 20:06:20","scene":"Shapes
    Scene","asset":"Button_memasangkan bentuk","events":"Player goes to match
    shapes game 1","task":""},{"time":"2023-05-12 20:06:22","scene":"Matching
    Shape game 1","asset":"jam","events":"Player is dragging
    asset","task":""},{"time":"2023-05-12 20:06:22","scene":"Matching Shape game
    1","asset":"jam","events":"Player is dragging
    asset","task":""},{"time":"2023-05-12 20:08:19","scene":"Main
    Scene","asset":"Button_keluar","events":"Player goes to closing
    scene","task":""},{"time":"2023-05-12 20:08:21","scene":"Closing
    Scene","asset":"exit button","events":"Game ends","task":""}]}
```


Figure 17: Example of games logs in JSON

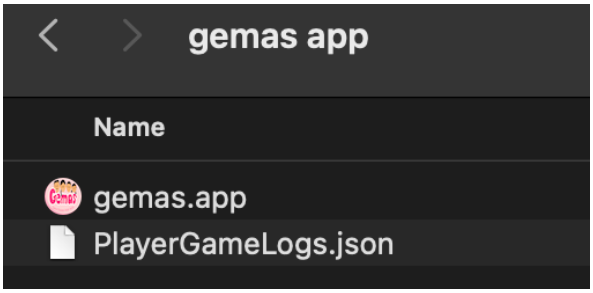


Figure 18: Gemas game logs is generated in the same path as the Game

4.3 GVT UI

In this section, the GVT final UI is presented. It was implemented based on the mock-ups presented in the section 3, with some designs were under modification to fit GVT functionalities.

GVT Desktop App

By extending the design, there are two more entry columns in GVT desktop app. Running on's entry column will show the host address when the Get button is clicked. This trigger the ED End Point's entry column to show the end point to input in the ED configuration in order for ED to be able to send the emotion logs into GVT server. As discussed in the section 3, Player Name entry is needed to input player's name. Game Folder entry uses to locate the game and game players logs path. Start button will commence the run of all GVT functionalities, while the stop button will stop all the GVT process. The realization of this design is shown in Figure 19.

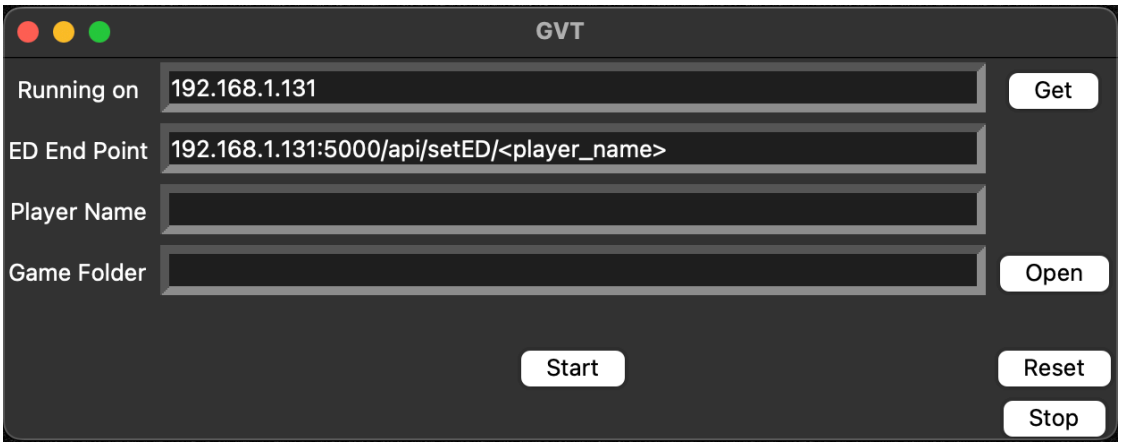


Figure 19: GVT desktop app UI

GVT Web App

As mentioned in the previous chapter, GVT has four main menus: Home, Single Analysis, Multi Analysis, and About. Following the original mock-up, the realization of **Home** page design can bee seen in the Figure 20.

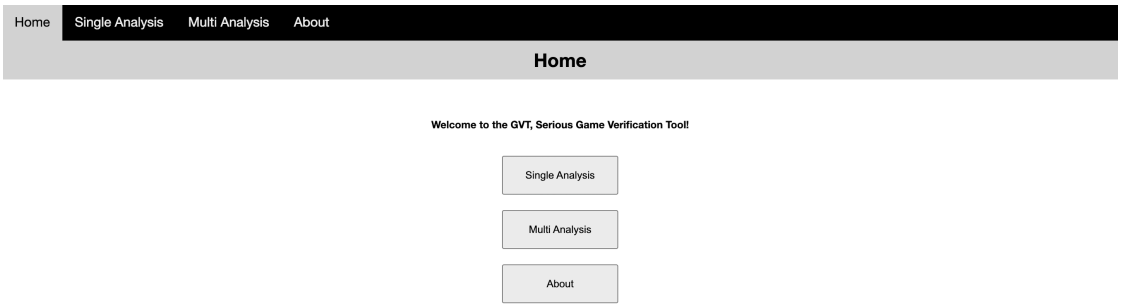


Figure 20: Home page UI of GVT web app

The second main menu is **Single Analysis**. The implementation of the design can be seen in Figure 21 with the logic of the select drop down options for Player Name and Date Play are depicted in Figure 22 and 23. Rational behind this function is that to accommodate the possibility of one player do the game testing more than once in separate days.

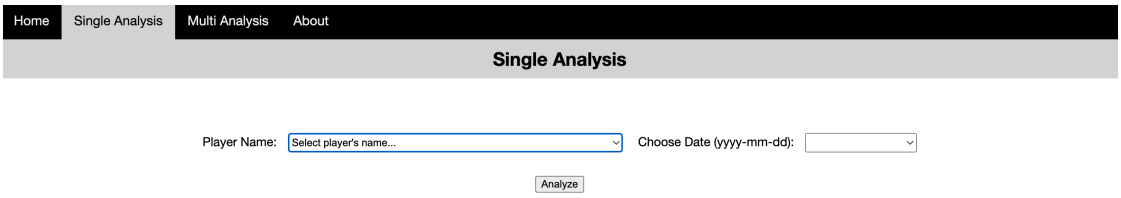


Figure 21: Single Analysis page UI of GVT web app

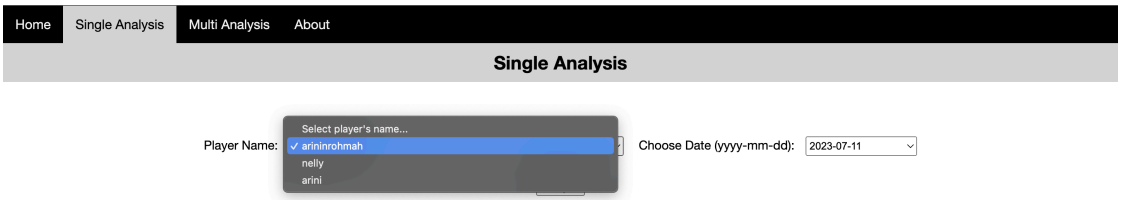


Figure 22: Single Analysis page Player Name function UI of GVT web app

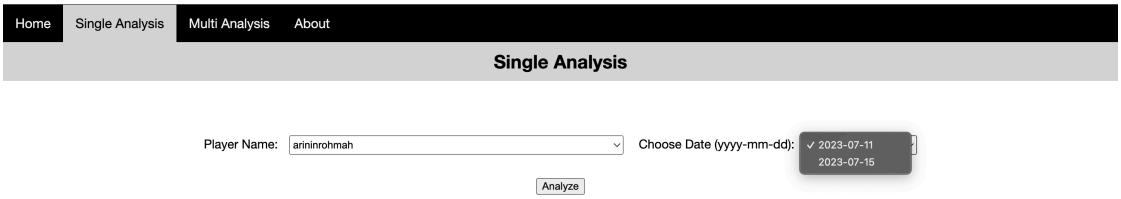


Figure 23: Single Analysis page Date Play function UI of GVT web app

Following that, after analyze button is clicked, GVT will show the analysis result which includes tables and graph. There are three tables in result of the analysis process: Monitoring Data, Single Analysis, Analysis Based on Scene. Additionally, there is a Glossary table to help GVT users understand the formula used in the analysis process. Further, based on the Analysis Based on Scene table, dynamic graph will be presented.

Figure 24 shows the Monitoring Data table. The table is intended to show the raw records of merging the emotion logs and game player logs. The Monitoring Data table has six columns

which are Time, Scene, Asset, Event, Task (completion), and Stress level. Each of the column represents the records accordingly from the selected player records in database. The table can be downloaded as .csv file through the download button.

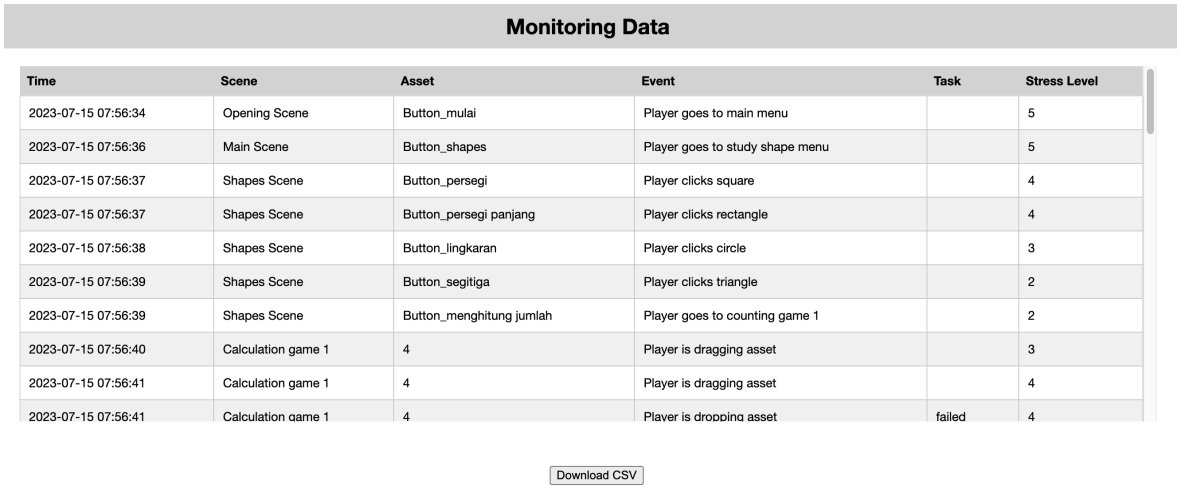


Figure 24: Single Analysis page, Monitoring Data table UI of GVT web app

The Single Analysis table is depicted in Figure 25. Through this table, GVT shows the first analysis process which are calculating Time Spent and time Accumulation. Thus, there are two new columns to accommodate Time Spent and Time Accumulation. The table is also colored based on the stress level. From (Suni Lopez et al. 2019), stress level was categorized into three ranges: Low, Medium, High. Low, if the stress level value less than 3. Medium, if the stress level value between 3 and less than 4. High, if the stress level equal or more than 4. Those categorisations are represented through colors, in which Green represents Low, Yellow denotes Medium, and Red indicates High stress level. Furthermore, this table provides sorting functions in each column, as well as filter to filter the table based on the Task completion. Figure 26 shows the example of the table when the filter is active. Single Analysis table can be downloaded as a .csv when the download button is clicked.

Figure 25: Single Analysis page, Single Analysis table UI of GVT web app

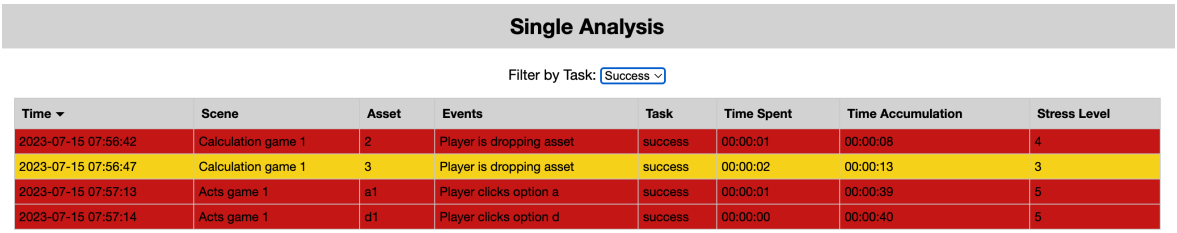
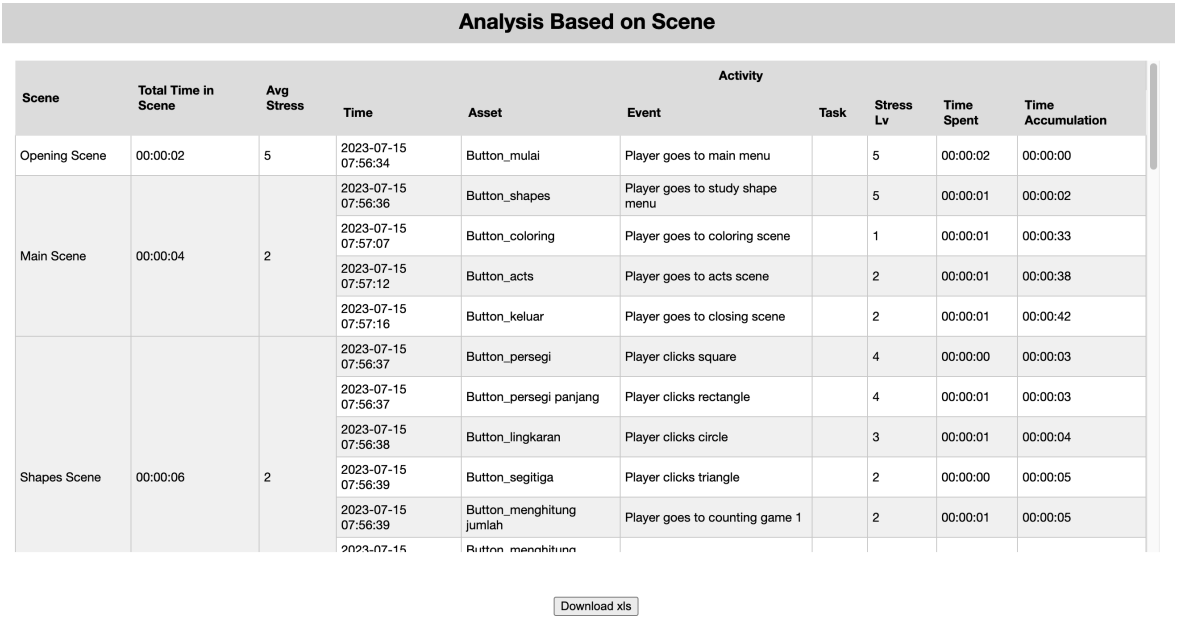


Figure 26: Single Analysis page, Single Analysis table task filter UI of GVT web app

Analysis Based on Scene table is designated the records based on scene. The example of the table is presented in Figure 27. From the raw records, GVT groups them accordingly into the same scene when the events are recorded. It is then calculated to generate the Average Stress Level for each scene, along with the Total Time spent in that particular scene. In result, the table expands with two more columns: Total Time in Scene and Avg Stress. Using this table, developers can start to analyse which scenes that have high average stress level. Therefore, they can indicate in which activities in that particular scenes that make the average stress level score high. In result, GVT can encourage serious game developers to modify those elements to improve the game for lowering the stress level score. Additionally, the table can be downloaded as an .xls file by clicking the download button.



Analysis Based on Scene

Scene	Total Time in Scene	Avg Stress	Activity						
			Time	Asset	Event	Task	Stress Lv	Time Spent	Time Accumulation
Opening Scene	00:00:02	5	2023-07-15 07:56:34	Button_mulai	Player goes to main menu		5	00:00:02	00:00:00
Main Scene	00:00:04	2	2023-07-15 07:56:36	Button_shapes	Player goes to study shape menu		5	00:00:01	00:00:02
			2023-07-15 07:57:07	Button_coloring	Player goes to coloring scene		1	00:00:01	00:00:33
			2023-07-15 07:57:12	Button_acts	Player goes to acts scene		2	00:00:01	00:00:38
			2023-07-15 07:57:16	Button_keluar	Player goes to closing scene		2	00:00:01	00:00:42
Shapes Scene	00:00:06	2	2023-07-15 07:56:37	Button_persegi	Player clicks square		4	00:00:00	00:00:03
			2023-07-15 07:56:37	Button_persegi panjang	Player clicks rectangle		4	00:00:01	00:00:03
			2023-07-15 07:56:38	Button_lingkaran	Player clicks circle		3	00:00:01	00:00:04
			2023-07-15 07:56:39	Button_segitiga	Player clicks triangle		2	00:00:00	00:00:05
			2023-07-15 07:56:39	Button_menghitung jumlah	Player goes to counting game 1		2	00:00:01	00:00:05

Download xls

Figure 27: Single Analysis page, Analysis Based on Scene table UI of GVT web app

GVT uses graph to provide its users with the simple and concise visualisation of the analysed data. There is a dynamic graph that can be changed accordingly to generate a comparable combination between Scene, Stress level, and Total Time. The graph has two types that can be selected: Line and Bar. Users are expected to input the combination as their desire and when the Generate Chart button is clicked, the graph will show the visualisation accordingly. Figure 28 and 29 depict the example of generated graph. The first aforementioned graph shows the combination of Stress Level by Scene as a line graph, while the last shows a bar graph of Total Time by Scene.

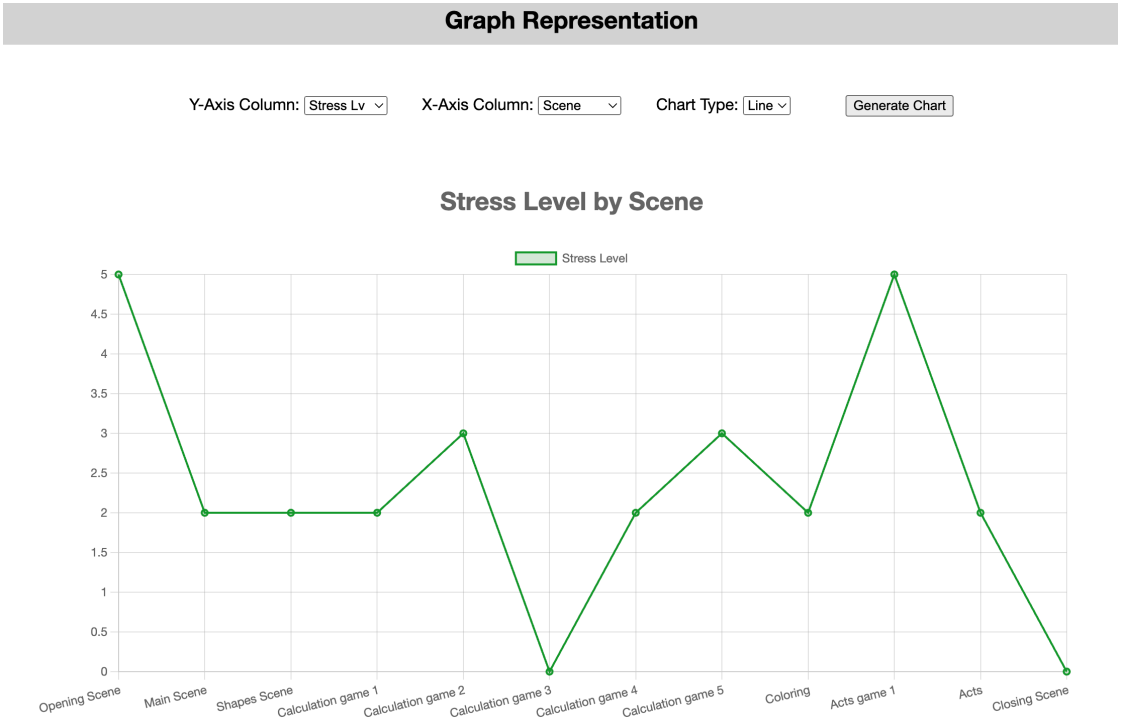


Figure 28: Single Analysis page, Stress Level by Scene graph UI of GVT web app

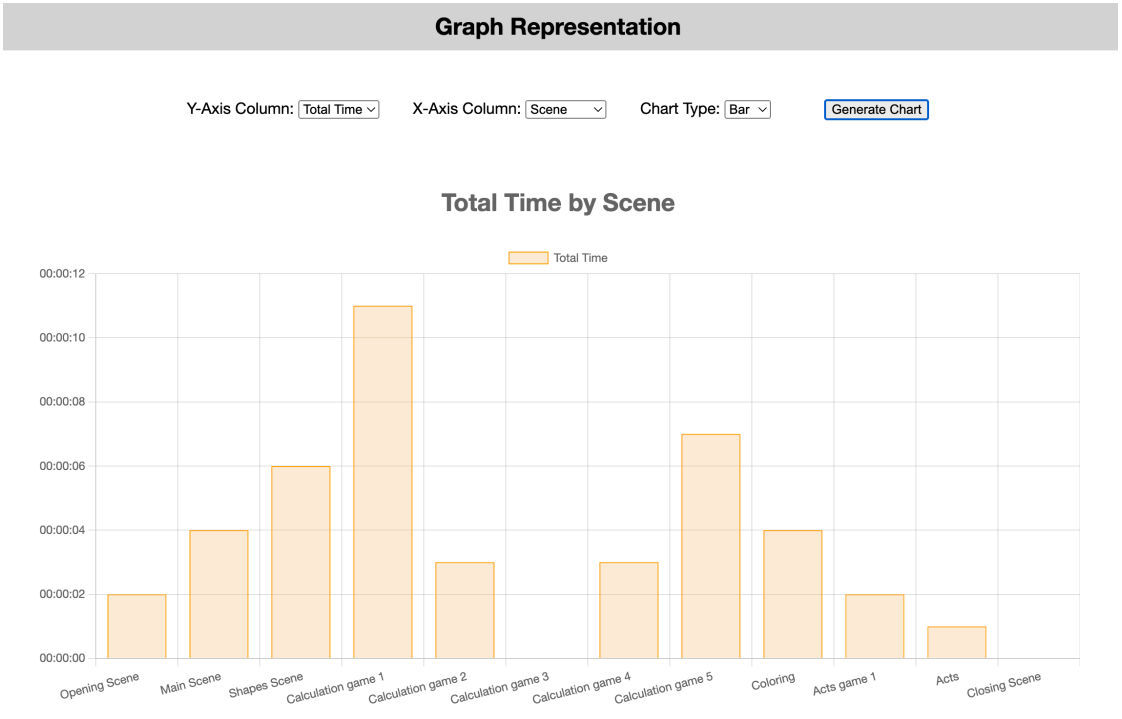


Figure 29: Single Analysis page, Total Time by Scene graph UI of GVT web app

The last table generated in the Single Analysis page is Glossary table. As can be seen in Figure 30, this table is provided to be useful for the GVT users if they want to know what formula is used to process the data records. Four formulas are presented in the table: Time Spent, Time Accumulation, Total Time in Scene, and Average Stress Level. Those of that are the main formulas used to calculate and analyse data records which the values are generated inside the aforementioned tables (Single Analysis and Analysis Based on Scene).

Glossary

Term	Definition	Formula
Time Spent	Represents the time elapsed between the current event and the next event in the game. It is calculated by finding the next event time and subtracting the current event time from it.	time_spent = (next_event_time - event_time).total_seconds()
Time Accumulation	Represents the time elapsed since the first event until the current event. It is calculated by subtracting the previous event time from the current event time, and adding the result to the total time accumulation.	time_accumulation = time_accumulation + (event_time - prev_event_time).total_seconds()
Total Time in Scene	Represents the total time spent in a particular scene. It is calculated by adding the time spent on each activity in that scene.	total_time_in_scene = sum(activity['time_spent'] for activity in data[scene]['activities'])
Avg Stress	Represents the the average stress when player play in the scene.	the sum of ((stress * time spent) in each activity) devided by total_time_in_scene

Figure 30: Single Analysis page, Glossary table UI of GVT web app

The third main page is **Multi Analysis** page. Figure 31 visualizes the realization of the design. Button Add allows users to add more players records to be selected, while the small red button on the right side of the form allows users to remove the selected players logs to not be included into the analysis. Further, Analyze button initiates the analysis process when it is being clicked.

In result of the initiations of Analyze button, GVT generates assessment Result of the feasibility score of the tested serious game. It gives Overall Stress Level score along with Overall Time Play. The final score is from the Overall Stress Level score whis is categorized into three categorizations: Stressless, Neutral, and Stressful. The same clusters from (Suni Lopez et al. 2019) are applied. The score is considered Stressless if the value less than 3. Neutral, if the overall stress level value between 3 and less than 4. Stressful, if the stress level equal or more than 4. Additionally, GVT presents the Stress Level Distribution chart that is showing the ratio of stress level from all the records. Using this information, game developers will directly get the summary of the assessment process of multi-records analysis.

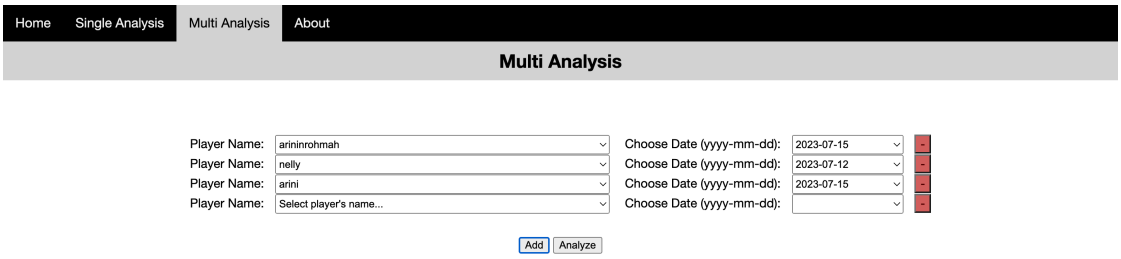


Figure 31: Multi Analysis page UI of GVT web app

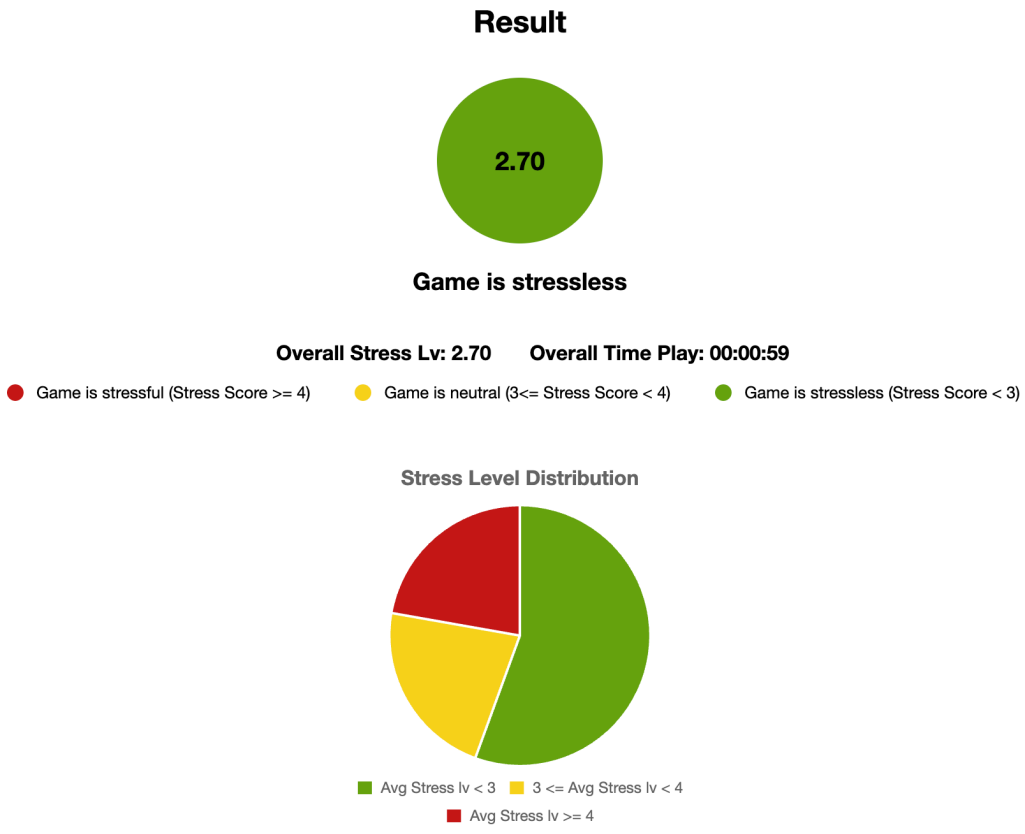


Figure 32: Multi Analysis page, game feasibility Result UI of GVT web app

To give more details, GVT also provides Result Detail table as seen in Figure 33. GVT calculates all the multi-records by extending the logic of formulas used for calculating single-records. Furthermore, the formulas allow GVT to calculate the Overall Stress Level and Overall Time Play.

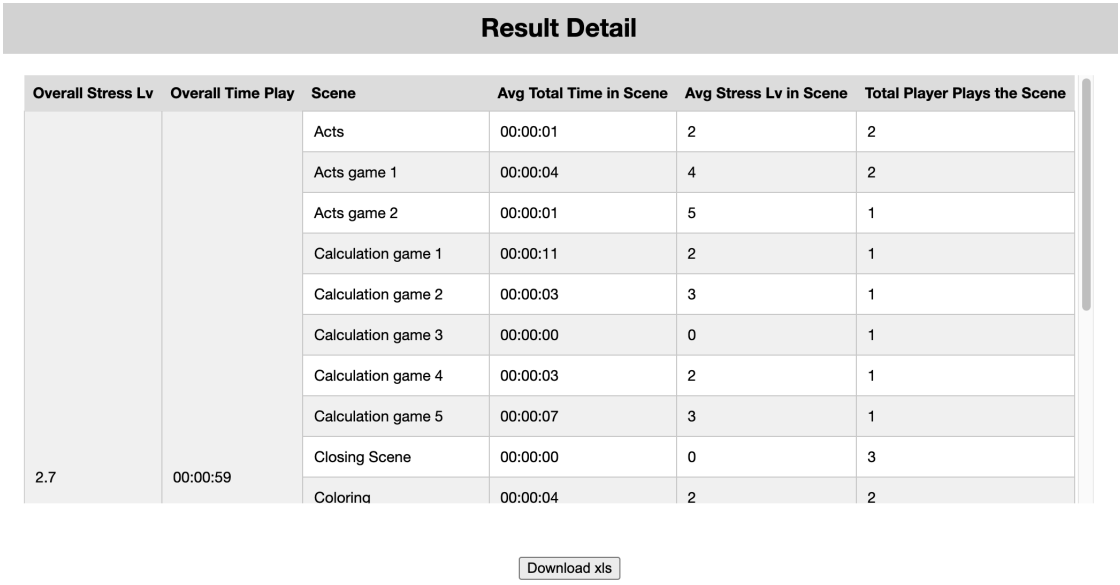


Figure 33: Multi Analysis page, Result Detail table UI of GVT web app

Based on the calculated records that are shown in the Result Detail table, a dynamic Result Graph is generated. It is shown the final calculation of multi-records (which is turned into single-records data) into a chart. As mentioned before, the chart is dynamic, the users are allowed to configure the input before generating desired chart. There will be a combination between Avg Stress Level, Avg Total Time, and Scene, along with the type of the chart. For instance, Figure 34 shows the Line chart of Avg Stress Level by Scene, while Figure 35 shows the Bar chart of Avg Total Time by Scene.

Additionally, there is a possibility that the games developers want to compare each player records. Therefore, Multi Analysis page accommodates this concern by providing Detailed Graph Representation - Multi Players. This graph visualizes comparable multi records point into one graph. Having the same functionality, the users can generate a graph as their desired. Figure 36 and 37 represent the detailed graph of multi records as a line chart and bar chart, respectively. While Figure 36 shows the Avg Stress Levels by Scene, Figure 37 presents the Avg Total Time by Scene.

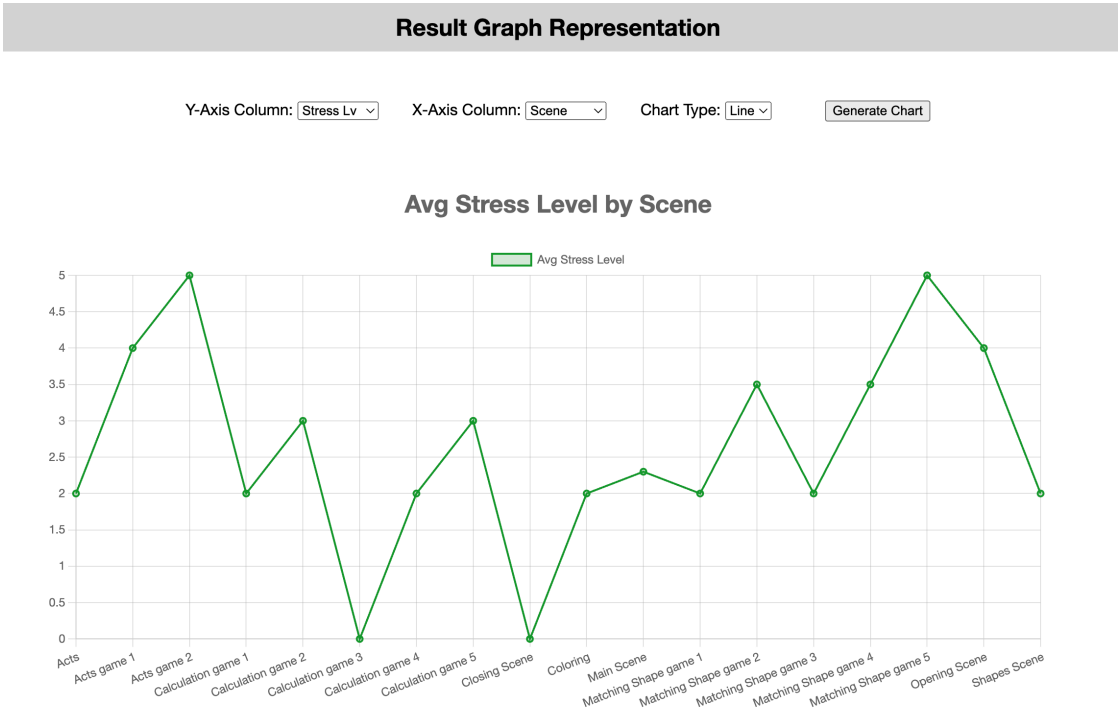


Figure 34: Multi Analysis page, Line graph of Avg Stress Level by Scene of Result UI of GVT web app

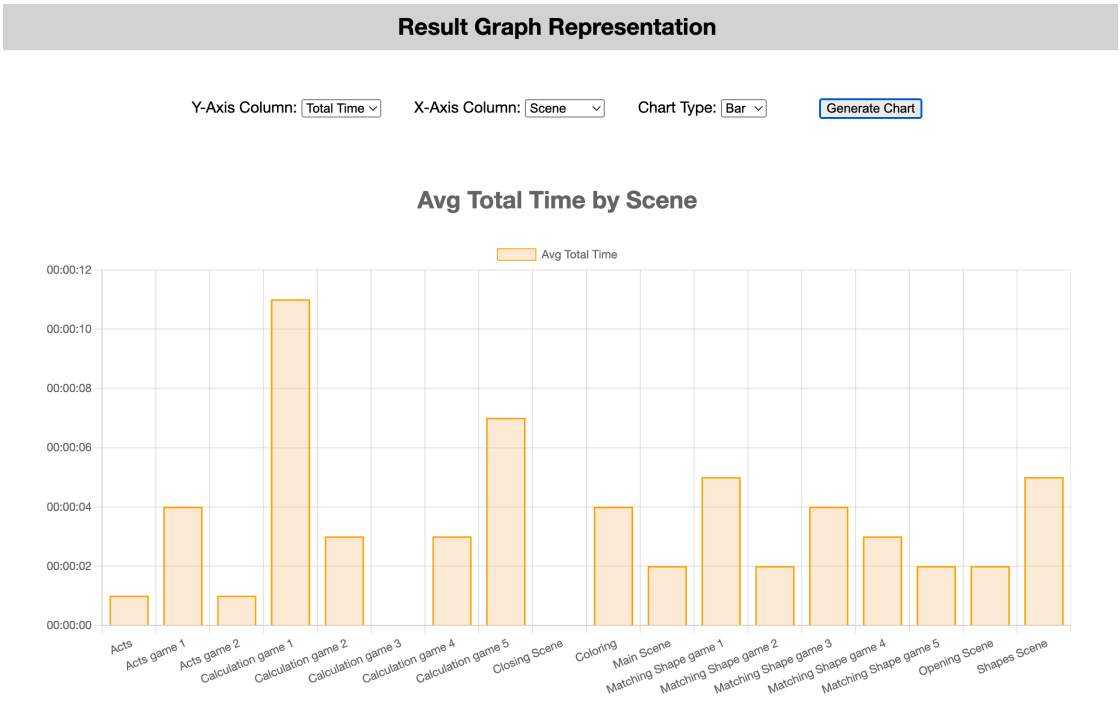


Figure 35: Multi Analysis page, Bar graph of Avg Stress Level by Scene of Result UI of GVT web app

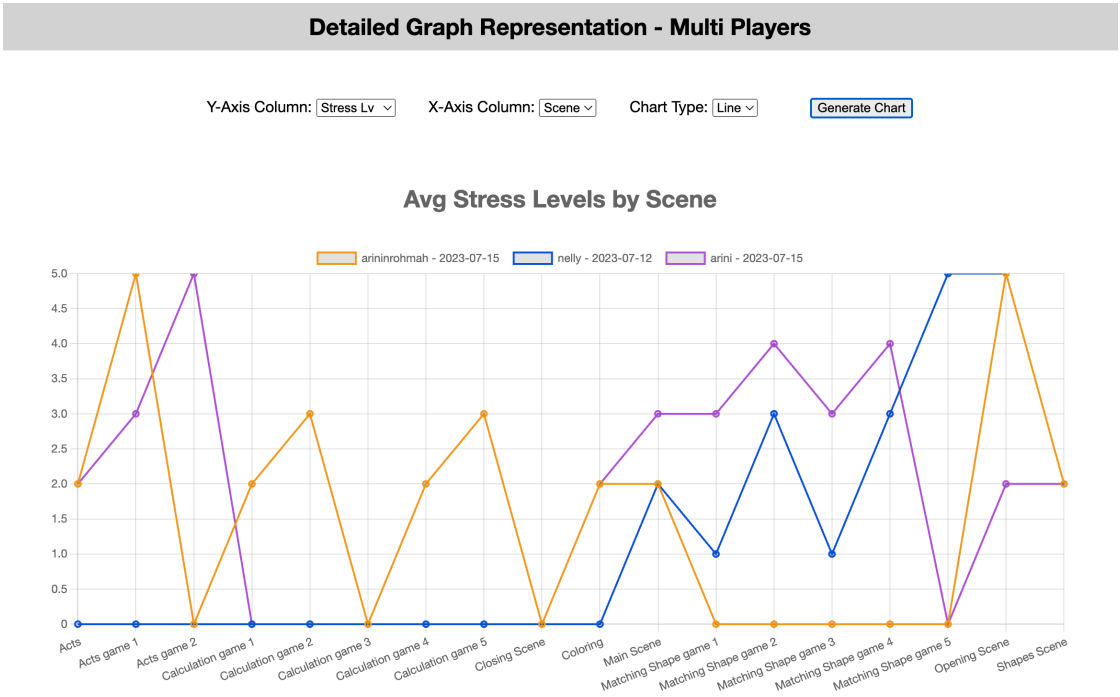


Figure 36: Multi Analysis page, Line graph of Avg Stress Lv by Scene of multi players records UI of GVT web app

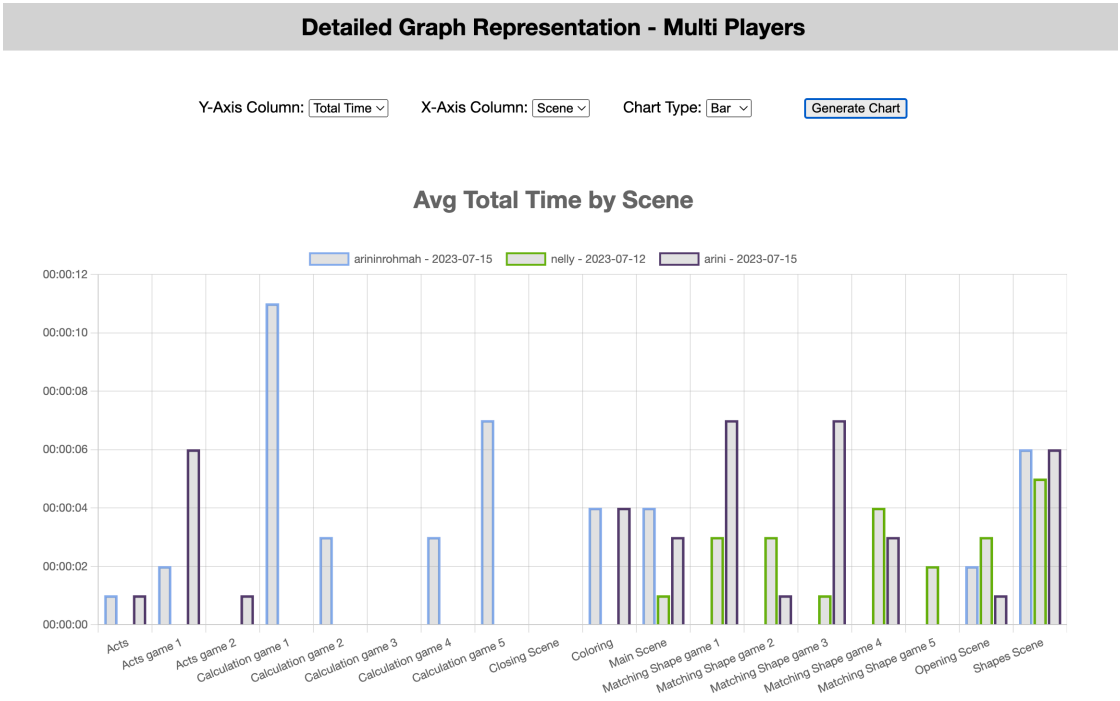


Figure 37: Multi Analysis page, Bar graph of Avg Total Time by Scene of multi players records UI of GVT web app

The Glossary table is the last table generated in the Multi Analysis page. As can be seen in Figure 38, this table shows the four extending formulas used to calculate the Average Stress Level Each Scene, Average Total Time Each Scene, Overall Stress Level, and Overall Time Play. Those of that formulas are the main formulas used to calculate and analyse multi data records which the values are generated inside the Result Detail table. Glossary table is intended to be useful for the GVT users if they want to know what formula is used to process the data records.

Glossary

Term	Definition	Formula
Avg Stress Level Each Scene	Represents the average stress level in each particular scene. It is calculated by summing the multiplication of the all average stress and average total time of that particular scene divided by the total number of records in that particular scene.	avg_stress_scene = (sum(avg_stress_scene * avg_total_time_scene))/total_records
Avg Total Time Each Scene	Represents the duration of players play in each particular scene. It is calculated by summing all the average total time in scene divided by the total records that particular scene.	avg_total_time = (sum(avg_total_time_scene))/total_records
Overall Stress Level	Represents the overall stress level of the game.	overall_stress = (sum(avg_stress_scene * avg_total_time)/total_scene)
Overall Time Play	Represents the overall time play of the game.	overall_time_play = (sum(avg_total_time))

Figure 38: Multi Analysis page, Glossary table UI of GVT web app

About page gives the information about the GVT. The realization of the About page design can be seen in Figure 39.

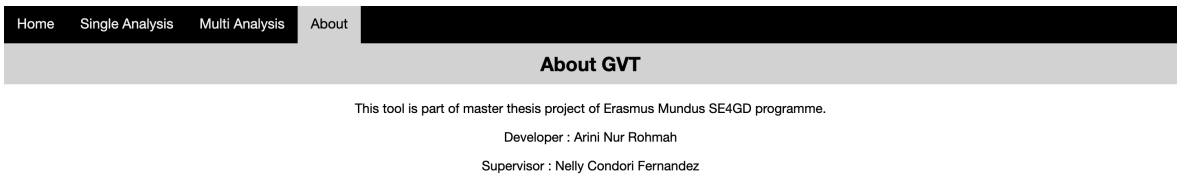


Figure 39: About page UI of GVT web app

5 TESTING AND EVALUATION

This chapter will discuss about the testing and evaluation process of GVT. The GVT usability test was conducted in CiTIUS which was involving in total of 14 participants which are developers and programmers. Note that the testing was intended to test the usability of the GVT and not for testing games. Therefore, only comments and advice from developers or programmers point of view are needed. Thus, there were no autistic children involved in the testing and evaluation process. Prior to the testing process, there was a need to select the emotion detector (ED) and the serious games (SG) targeting autistic children used for testing the GVT. Those of that along with the testing process and evaluation will be elaborated in separate sections.

5.1 Selection of Emotional Detector and Serious Game

It was necessary for GVT to find a suitable baseline of ED and SG used for testing. As mentioned in the Chapter 3, in collaboration with CiTIUS, the ED used in the testing process was an adapted version of a mobile app developed by Mamani 2021, which combines EDA signal and speech in english. For the usability testing, the mobile app was modified, by considering only the EDA signal as input of the stress detector and allowing the persistence of the detector outcome that was sent to the server. The real-time stress detector, using EDA signal, implements the pipeline proposed by Suni Lopez et al. 2019. Although underwent some modifications, the accuracy of the stress detector does not change. As reported in Mamani 2021, ED has an accuracy of 73.5% in real-time conditions.

Figure 40: ED used during usability testing: EDA sensor (Left) and adapted version of a mobile app (Right), Mamani 2021 (ibid.)

Figure 40 shows how the sensor and ED look like. As can be seen in that figure, players are expected to wear wearable sensor which are worn in two fingers of non dominant hand. The sensor then sends the raw data of emotion into the ED mobile device through BT. ED then analyzes the emotion data, then sends the emotion (stress level) logs to the GVT server end-point via local host. Meanwhile, for selecting the SG, an existing serious game targeting autistic children, called Gemas (Al Irsyadi and Rohmah 2017) was modified for this purpose. The learning sub-

ject materials that are taught by Gemas are based on a textbook used by Rumah Pintar Salatiga, a special school for people with special needs in Indonesia. Gemas, initially was a static game for autistic children based on Kinect which lacked the ability to generate player logs. Also, Gemas only in Indonesian language which made it inaccessible for Spanish players. Subsequently, Gemas underwent modifications and transformed into a desktop game capable of generating player logs as well as having Spanish language as a new language option. The modification process took nearly a month. The detailed description of Gemas is as follows: **Goal:** To assist teachers in introducing shapes, colors, numbers, and examples of good behavior to school-age children with autism in an engaging and enjoyable manner, aiming to enhance their interest in learning these subjects at school. **Target:** School-age children with autism. **Platform:** Desktop platform. **Objective:** Implementation within school settings. **Language:** Spanish and Indonesian language. **Technical info:** The game consists of a total of 20 scenes and is considered static in terms of difficulty level and reinforcement feedback. The hierarchy of Gemas' scenes can be seen in Figure 41.

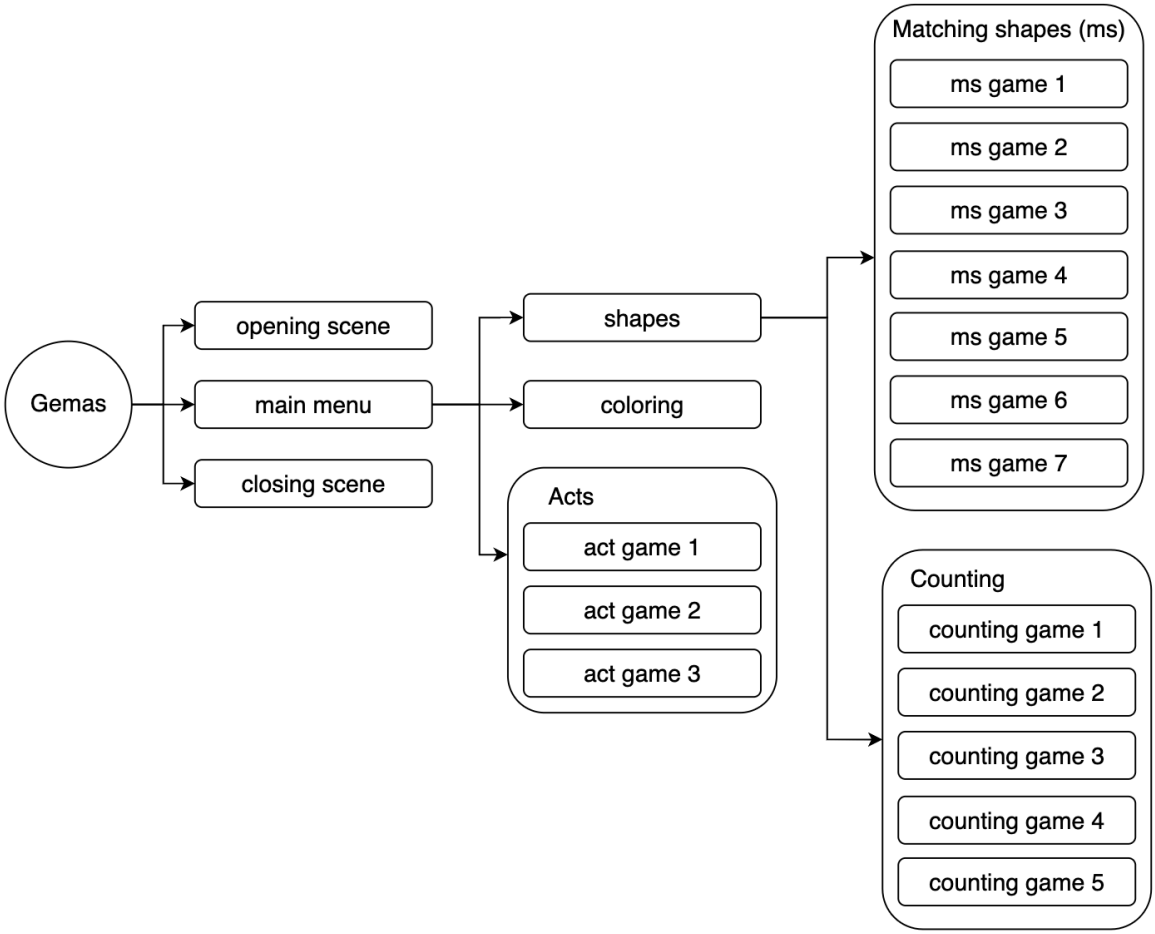


Figure 41: Gemas game hierarchy

5.2 Usability Evaluation Design

The plan and design of the GVT usability evaluation has been carried out according to the usability evaluation method proposed by Condori-Fernández et al. 2013.

5.2.1 Subjects and Setting

There were in total 14 participants involved in usability evaluation of the GVT. The participants are developers and programmers which one of them is visually impaired programmer. The testing process was held in CiTIUS in the same room and conditions for approximately three days during the day. Author acted as a SG player in order to obtain the emotion and game player logs for the GVT to analyze in the testing process.

The example setting of the testing process can be seen in Figure 42a and 42b. Figure 42a shows how Author played the SG after the participants ran the GVT desktop app in order to obtain game player records. Meanwhile, Figure 42b shows the example on how the developers or programmers who participated in the usability testing of the GVT web app did the evaluation.

(a) how expected player do the test

(b) how expected SG developers analyze their games using GVT

Figure 42: GVT usability test

Variables

The usability is evaluated based on three quality aspects: Efficiency, Effectiveness, and User Satisfaction. According to Condori-Fernandez and Lago 2018, **efficiency** is defined as "resources expended in relation to the accuracy and completeness with which users achieve goals". **Effectiveness** means "accuracy and completeness with which users achieve specified goals". While **user satisfaction** is related to users confidence that the system or product will behaves as desired and when the goals are perceived, it satisfies the users. Meanwhile, according to Chung et al. 2012, efficiency is related to the performance, in which how well the

product or system is functioning in utilizing a resource. In correlation to that, effectiveness is one of the concern in the performance efficiency.

Usability Tasks

Participants were given list of tasks related to the GVT functionalities to test. The list of tasks and corresponding activities (the full list can be seen in Appendix B) is as follows:

- Desktop app
  - T1 Configure GVT
    - A1.1 Modify GVT entries
    - A1.2 Starting GVT
    - A1.3 Stopping GVT
- Web App
  - T2 Analyze Single data record from Single Analysis page
    - A2.1 Monitoring single player's records
    - A2.2 Download Monitoring Data table
    - A2.3 Analyse single player's records
    - A2.4 Filter and sort Single Analysis Table
    - A2.5 Download Single Analysis table
    - A2.6 Grouping the records based on scene
    - A2.7 Download records based on scene
    - A2.8 Show the analysis as a Graph
    - A2.9 Read the Glossary for single analysis
  - T3 Analyze multi data records from Multi Analysis page
    - A3.1 Do Multi Analysis
    - A3.2 Show the result as a table
    - A3.3 Download Result Detail table
    - A3.4 Show the Result Detail table as a graph
    - A3.5 Compare the multi user's records in a chart
    - A3.6 Read the Glossary for multi analysis

All of the activities in each task were defined by considering all the functionalities the GVT has. The activities used to identify the completeness of the tasks. It then helped to formulate the questionnaire in the evaluation process.

Instrumentation

There were three **instruments** used for the usability test, they are as follows:

- Recorder (screen and voice): to time the users as well as to capture and identify tasks completion in regard for efficiency evaluation.

- List of tasks: tasks to carry out by participants related to GVT functionalities.
- Post-test questionnaire: to get full feedback from the participants in relation with effectiveness and users satisfaction.

The **recordings** was used to time the users as well as to capture and identify tasks completion for evaluating **efficiency**. Task is considered complete is the user complete all the activities specified in the appendix B. It is calculated from the average total time spent from starting until finish the evaluation. Besides, Thinking aloud technique (Waes 2000) was adopted to encourage participants express their mind. Thus, screen and voice recorder was utilized to tape participants voice and their screen activities.

Post testing process, all the participants were expected to fill Google Form **questionnaire** and give their objective opinions. Google Form was mainly (but not particularly) used to evaluate the effectiveness and the user satisfaction. There are in total 39 questions, which consist of 34 required questions and 5 optional questions. The optional questions are required to gather qualitative feedback related to comments and suggestions from participants. **Effectiveness** was calculated in two measurements criteria: manageable and unmanageable, by using a linear scale in regards from the Q2 (see Table 13) answer. In correlation with effectiveness, the tasks were classified into three categories: simple, medium, complex. Meanwhile, the **user satisfaction** was calculated using nonverbal measurement tools - Emocards (Desmet 2000; Agarwal and Meyer 2009) which having eight emotion scales (as seen in Figure 43) in order to be more dynamic and precise to express the users emotion. Additionally, user satisfaction includes how user is satisfied with their perceived achievement of pragmatic goals (Condori-Fernandez and Lago 2018). In the GVT case, one of the most important **pragmatic goals** is that the analysis reports shown in the GVT web app are understandable, therefore, linear scale questions related to this aspect (Q1) are included inside the Google Form sheet.

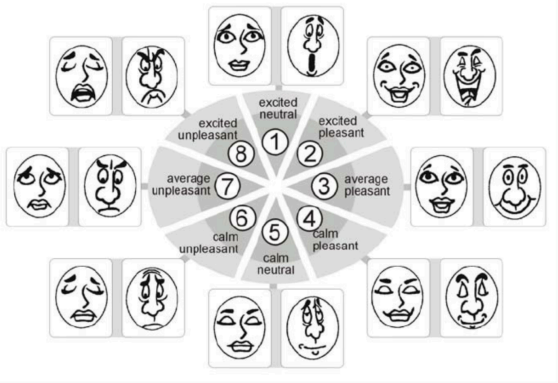


Figure 43: Emocards used for evaluating users' satisfaction

There are two main sections in questionnaire related evaluating the usability of GVT: GVT desktop app and GVT web app. For evaluating GVT desktop app, there are only two questions and one entry suggestion in the questionnaire. Meanwhile, for evaluating GVT web app, there are four separate sections in relation with the menus: Home, Single Analysis, Multi Analysis, and About. The list of full questions for both GVT desktop and web app can be found in Appendix C.

Each functionality consists of two questions. However, some of the functionalities have three questions related to the effectiveness and satisfaction according to their function. Table 13 shows the main questions for evaluation. The only difference between Q1 and Q2 is in the user decision needed to run the function. Q1 was intended for both the functionalities which require any decision from the user or not, while Q2 was intended only for functionalities which require user decision.

Table 13: Question related to Effectiveness (Q2) and Satisfaction (Q1 and Q3).

Symbol	Question	Scale
Q1	Is it understandable?	5 linear scale with 5 is the most understandable
Q2	Is it manageable?	5 linear scale with 5 is the most manageable
Q3	Are you satisfied?	8 scale

The **tasks** were classified into three categories: simple, medium, complex. The classifications of the tasks are as follows:

- Medium: T1
- Complex: T2 and T3

Meanwhile, the details of **devices** used for the usability testing are as follows:

- MacBook with operating system Ventura 13.4.1. This device used by participants to test the usability of the GVT desktop app and to run the SG played by Author who acted as a game player.
- iPad with operation system iPadOS version 16.5.1. This device used by participants to test the usability of GVT web app.

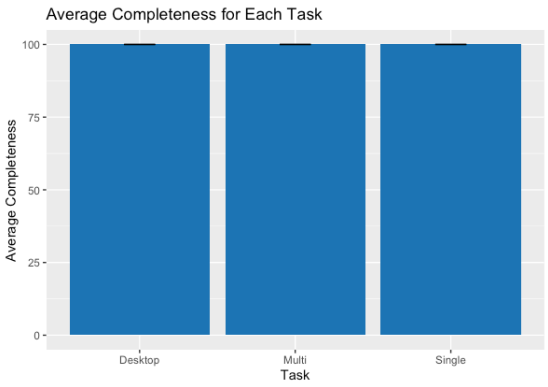
5.3 Usability Evaluation Process

There are two evaluation processes based on the methods gathered from the usability test: **recordings** and **questionnaire**. 14 recordings (screen and voice) as well as 14 filled questionnaires were collected from the usability test.

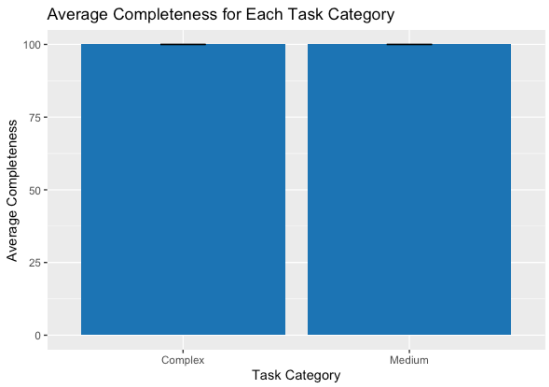
5.3.1 Data Analysis and Results

The information gathered from video and voice recordings shows that all of the participants complete the tasks fully, therefore the **completeness** is 100% as can be seen in Figure 44(a) and 44(b) for completeness each task and task categories, respectively. The testing result was plotted into bar graphs according to task and task categories for efficiency, effectiveness

and users satisfaction. The graphs call task 1 (T1), task 2 (T2), and task 3 (T3) as Desktop, Single, and Multi, respectively to represents GVT desktop app and web app menus.



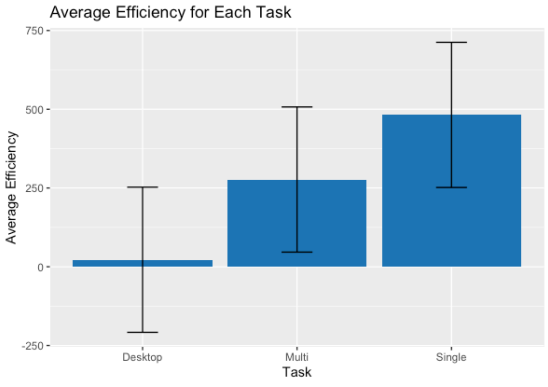
(a) Completeness of each task



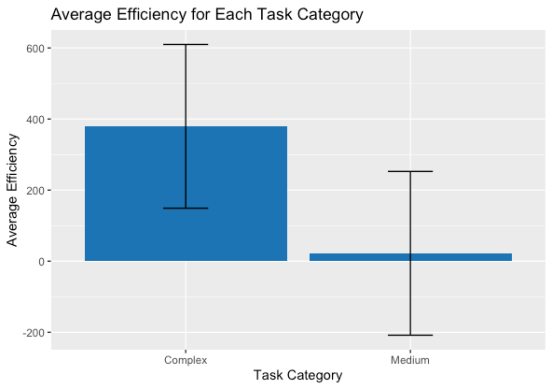
(b) Completeness of each task category

Figure 44: GVT Completeness

Figure 45(a) represents the **efficiency** of GVT in relation of average time spent of tasks (in second) and tasks while Figure 45(b) shows the relation of average time spent of tasks (in second) with difficulty categories. As can be seen from the figures, the complexity of tasks affect the time spent on completing the tasks.



(a) Average efficiency of each task



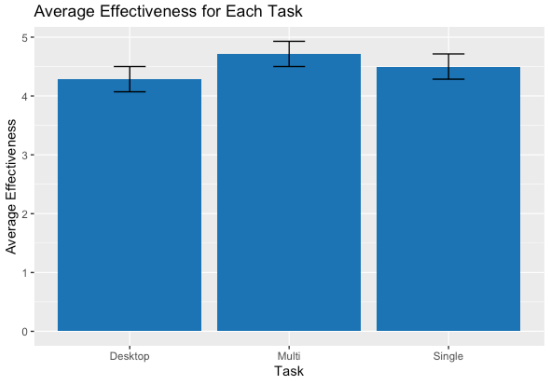
(b) Average efficiency of each task category

Figure 45: GVT efficiency

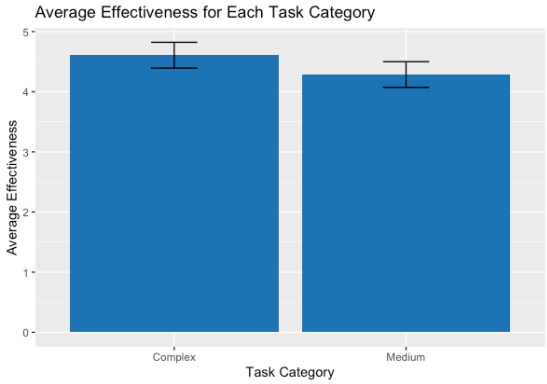
Since the completeness of tasks all gained 100%, for effectiveness, we consider to depict data from the questionnaire. According to the questionnaire filled by participants along with the recordings, the average score related to the management tasks are high. All of the tasks are considered manageable with the score in range 4-5 (highest scale) as can be seen in Figure

47(a). However, medium tasks got the lowest score of **effectiveness** compared to complex tasks. This odd occurrence can be seen in Figure 47(b)

Based on the feedback, GVT satisfied participants with its functionalities. The **satisfaction** score can be seen in Figure 46(a) for average satisfaction by tasks and 46(b) for average satisfaction by task categories. Overall, all participants felt pleasant (range 2-4 of emocards) in using GVT with all the functionalities it has. Additionally, according to the complexity, participants felt the most satisfied when working on the complex tasks. Although it is quite strange that the complex tasks generate the best satisfaction score, according to the data received in the testing process, it is what the participants felt.

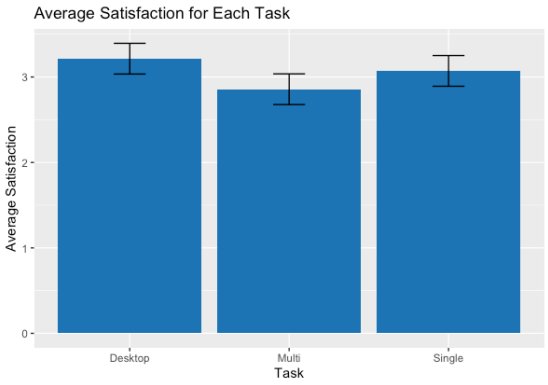


(a) Average effectiveness of each task

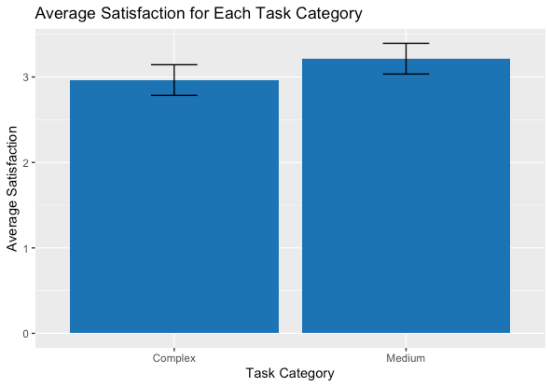


(b) Average effectiveness of each task category

Figure 46: GVT effectiveness



(a) Average satisfaction of each task



(b) Average satisfaction of each task category

Figure 47: GVT users satisfaction

Furthermore, the detailed evaluation of the results gathered from the questionnaire and recordings is elaborated in Table 14 and 15 for GVT desktop app and GVT web app, respectively.

Table 14 shows that the evaluation of **usability score of GVT desktop app** is relatively high. 50% of responders vote the highest score of **effectiveness** (5), along with the majority of **satisfaction score** is in average pleasant range (3). However, GVT needs to improve its UI to accommodate inclusiveness. In result of the lack accessibility from GVT desktop app to be an inclusive verification tool, 7.1% responders vote GVT as unmanageable app with satisfaction score in average unpleasant range.

As it was mentioned before, in addition of voice recording, the questionnaire provides one entry for suggestion in every sections for **qualitative feedback**. Related to GVT **desktop app**, there are 10 suggestions received. In summary, GVT design is good, efficient (but a bit bulky), easy to understand and manage, simple but usable. However, many aspects need to be improved, for instance tonality of colors and the size of letters. Furthermore, GVT desktop app is inaccessible for visually impaired users, or in general, for people with special needs. Ergo, improving GVT to be an inclusive verification tool is a must.

Table 14: Evaluation of GVT desktop app.

Total responders :	14	
Category		Result
Effectiveness	Q2	5: 50%, 4: 42%, 1: 7.1%
Satisfaction	Q3	3: 35.7%, 4: 28.6%, 2: 21.4%, 1: 7.1%, 7: 7.1%

Usability test evaluation for GVT web app is relatively high as can be seen in Table 15. Overall, more than 60% responders agreed that the **effectiveness** of the GVT is 5 out of 5 scale. Meanwhile, the majority of **users satisfaction** is reported to be in excited pleasant - average pleasant range. However, as it was stated before, GVT is still lack of accessibility interface. Thus, GVT is considered ininclusive for disabled users. Ergo, resulting in 7.1% of responders felt unsatisfied with GVT with the emotion scale is in average unpleasant - excited unpleasant.

Table 15: Overall evaluation of GVT web app.

Total responders :	14		
Menu	Category		Result
Home	Satisfaction	Q3	3: 42.9%, 2: 28.6%, 4: 14.3%, 5: 7.1%, 8: 7.1%
Single Analysis	Effectiveness	Q2	5: 64.3%, 4: 28.6%, 2: 7.1%
	Satisfaction	Q3	2: 42.9%, 3: 42.9%, 5: 7.1%, 8: 7.1%
Multi Analysis	Effectiveness	Q2	5: 78.6%, 4: 14.3%, 3: 7.1%
	Satisfaction	Q3	2: 50%, 3: 21.4%, 1,4,5,7: 7.1%
About	Satisfaction	Q3	2: 28.6%, 1: 21.4%, 3: 21.4%, 4,5,7,8: 7.1%

By going into details to evaluate the **usability of Single and Multi Analysis** menus, Table 16 and 17 show elaborated evaluation according to all functionalities in Single and Multi Analysis menus, respectively.

Table 16: Evaluation of GVT web app - Single Analysis.

Total responders :	14		
Functionality	Category		Result
Monitoring Data Table	Satisfaction	Q1	5: 57.1%, 4: 42.9%
		Q3	2: 42.9%, 4: 28.6%, 3: 14.3%, 5,7: 7.1%
Single Analysis Table	Effectiveness	Q2	5: 71.4%, 4: 21.4%, 3: 7.1%
	Satisfaction	Q1	5: 64.3%, 4: 28.6%, 3: 7.1%
		Q3	2: 42.9%, 3: 21.4%, 4,5: 14.3%, 7: 7.1%
Analysis Based on Scene Table	Satisfaction	Q1	5: 64.3%, 4: 21.4%, 3,2: 7.1%
		Q3	3: 50%, 2: 21.4%, 1,5,6,7: 7.1%
Graph Representative	Effectiveness	Q2	5: 71.4%, 4: 28.6%
	Satisfaction	Q1	5: 71.4%, 4: 28.6%
		Q3	2: 42.9%, 3: 21.4%, 4: 14.3%, 1,6,7: 7.1%
Glossary	Satisfaction	Q1	5: 64.3%, 4: 21.4%, 3: 14.3%
		Q3	2: 35.7%, 3: 28.6%, 1: 14.3%, 4,5,8: 7.1%

Table 17: Evaluation of GVT web app - Multi Analysis.

Total responders :	14		
Functionality	Category		Result
Assessment Result	Satisfaction	Q1	5: 50%, 4: 35.7%, 3: 14.3%
		Q3	3: 28.6%, 2,4: 21.4%, 5: 14.3%, 1,7: 7.1%
Result Detail Table	Satisfaction	Q1	5: 71.4%, 3: 14.3%, 4,2: 7.1%
		Q3	2: 42.9%, 3: 28.6%, 1,4,5,8: 7.1%
Result Graph Representation	Effectiveness	Q2	5: 78.6%, 4: 21.4%
	Satisfaction	Q1	5: 78.6%, 4: 14.3%, 3: 7.1%
		Q3	2: 42.9%, 3,5: 21.4%, 1,8: 7.1%
Detailed Graph Representation - Multi Players	Effectiveness	Q2	5: 78.6%, 4: 21.4%
	Satisfaction	Q1	5: 71.4%, 4: 28.6%
		Q3	2: 50%, 1,3: 14.3%, 4,5,8: 7.1%
Glossary	Satisfaction	Q1	5: 71.4%, 4: 14.3%, 3: 14.3%
		Q3	3: 28.6%, 2,4: 21.4%, 1: 14.3%, 5,8: 7.1%

In addition of the **qualitative feedback received by voice recorder**, in relation with four menus available in GVT web app, the entry for suggestion was separately provided for each menu. There are 5 responses received for Home menu from the questionnaire. In summary, **Home** is considered simple, easy to understand and use. Meanwhile, only 1 response received from the questionnaire for **About** menu. In summary, About serves a necessary information that is expected from this kind of page. Not much text makes it quick to read and understand the idea of the work.

The qualitative feedback received for **Single Analysis** and **Multi Analysis** menus are pretty similar. There are 5 responses from questionnaire for both Single and Multi Analysis menus. Combining with the voice records, in summary, both pages have good interfaces and functions which easy to understand and follow. However, not only lack of accessibility, some part of the functionalities need to be improved in order to locate the important data easily and directly. Moreover, some functionalities are expected to be automatic when there is a change, instead of always clicking button to generate new desired info.

In summary, GVT is usable. Participants found the GVT to be manageable (effective) and efficient. The GVT is considered satisfying by the participants as they were perceiving the tool as understandable.

6 CONCLUSIONS

Serious games may be one of the most effective ways to help children with ASD doing learning activities. However, considering the traits of children with autistic in which having the lack of emotion maintainability, there is a need to validate the suitability of the serious games which targeting them. This is in order to ensure that the serious games targeting autistic children do not trigger the raise of their stress level.

Examination of previous validation methods that involve users (players and experts/non-experts) and cost factors was held along with this work to cluster and map those existing validation methods. In result, it is found that most of the existing methods are aiming for general games or SG, they require the involvement of end-user in the V&V process. However, those approaches have not considered the limitation of children with mental disorder which makes the V&V process relatively uninclusive. Additionally, in relation to this master thesis work, the most closest research to evaluate games which may be applicable for evaluating serious games targeting autistic children is GUR. However, GUR implements physiological responses evaluation which requires equipment in which those are considered intrusive for mental disorder children especially children with ASD.

Many emotion detection methods which can be integrate with the verification tool exist. However, taken into account traits of person with ASD, majority of those methods are not applicable for them. Children with ASD have this particular traits: hard in expressing emotion, anxious, easily meltdown. Considering those traits, self-Report measures of emotions may not be reliable if we expect children with ASD to fill the questionnaire or use pictorial methods. Implementing facial expression, gestures, and voice detection or analyzer, are not plausible as well. Further, utilizing physiological system measurements, may not be a best solution for this case, since almost all the physiological system measurements require a person to wear on-skin equipment, which is considered obstructive for autistic people.

Having goal to be a simple but useful verification tool for this purpose, an unintrusive serious games verification tool (GVT) is proposed. GVT was designed to identify serious games (targeting autistic children) elements in order to assist game developers improving their serious games and creating more inclusive serious games for children with ASD. GVT works by integrating emotion detector (ED) (which developed by Suni Lopez et al. 2019), with games players logs generated by the serious games. Using GVT technique, game developers can keep on track in which logs that the autistic players' stress level is considered high. Ergo, it will encourage the game developers to improve those of that elements to maintain their

players (autistic children) stress level in a safe range and make the games suitable for them. Further, if GVT is widely used for this purpose, GVT was foreseen to be part in enhancing the quality of education for children with ASD. Additionally, by providing the more suitable learning activities options for children with ASD, GVT reduces inequalities and provides inclusiveness.

The result of GVT usability test shows that according to 14 participants (programmers and developers), GVT is usable, manageable, and understandable. Participants believe that GVT can help serious game developers (who are targeting autistic children) to evaluate their serious game elements in a way that is safe to involve children with ASD in the evaluation process.

6.1 Limitation

The main goal of GVT is to serve as a simple game validation tool, relying on game player logs and an ED. However, some serious games for autistic children do not have player logs or the structure of the logs is not suitable for GVT, which makes it difficult for GVT to analyze them. Additionally, GVT only works for desktop games (since it is assumed that larger media can help children with ASD to concentrate better), while there are also valuable serious mobile games that require analysis. Along with that, regarding the proposed game logs, future research should explore in which extent the proposed script (which was developed to generate logs for Gemas) can be applied to other games including mobile games.

Another limitation is related on how data being analysed and presented by GVT. Since GVT only use average in data calculation, there is a concern that analyzed data presented by GVT may not be reliable enough to represent complex data analysis. GVT should consider adequate calculation such as quartile, max, min, and standard deviation to accommodate more complex data. In affect, the graphs presented by GVT should as well enhance by not only bar or line, but also histogram, kernel, or bubble according to data relations.

6.2 Recommendation

There are some areas that can be improved in this research work. To address this, Author recommend that in further GVT development, GVT should consider for not only involving autistic players' stress level, but also incorporating a wider range of emotions, such as happiness, anxiety, shock, anger, and more. By doing so, it will open a possibility to integrate GVT with other future wearable ED devices.

Additionally, there is an acknowledgement of the GVT limitations as it currently relies on externally generated game player logs. To overcome this, it would be highly advantageous if the GVT could automatically generate game player logs internally. This enhancement would

enable the evaluation of both games with and without player logs, making the evaluation process more comprehensive.

Furthermore, future research should focus on utilizing GVT to evaluate serious mobile games designed for autistic children. This could yield valuable insights and contribute to the improvement of gaming experiences tailored to this specific audience.

REFERENCES

(N.d.). URL: https://sdgs.un.org/goals.

(N.d.). URL: https://www.python.org/about/.

(N.d.). URL: https://www.json.org/json-en.html.

Agarwal, Anshu and Andrew Meyer (2009). "Beyond usability: evaluating emotional response as an integral part of the user experience". In: *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pp. 2919–2930.

Al Irsyadi, Fatah Yasin and Arini Nur Rohmah (2017). "Game Edukasi Bagi Anak Autis Bertema Anggota Keluarga Berbasis Kinect Xbox 360". In: *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer* 8.2, pp. 739–746.

Alarcon-Licona, Susana, Lian Loke, and Naseem Ahmadpour (2018). "From Autism Educators to Game Designers: Integrating Teaching Strategies into Game Design for Autism Education Support". In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. OzCHI '18. Melbourne, Australia: Association for Computing Machinery, pp. 58–62. ISBN: 9781450361880. DOI: 10.1145/3292147.3292208. URL: https://doi.org/10.1145/3292147.3292208.

Als, Heidelise, Edward Tronick, and T Berry Brazelton (1980). "Affective reciprocity and the development of autonomy: The study of a blind infant". In: *Journal of the American Academy of Child Psychiatry* 19.1, pp. 22–40.

Anolli, Luigi et al. (2010). "Emotions in serious games: From experience to assessment". In: *International Journal of Emerging Technologies in Learning (iJET)* 5.2010.

Bakhuys Roozeboom, Maartje, Gillian Visschedijk, and Esther Oprins (2017). "The effectiveness of three serious games measuring generic learning features". In: *British journal of educational technology* 48.1, pp. 83–100.

Berkovits, Lauren, Abbey Eisenhower, and Jan Blacher (2016). "Emotion regulation in young children with autism spectrum disorders". In: *Journal of Autism and Developmental Disorders* 47.1, pp. 68–79. DOI: 10.1007/s10803-016-2922-2.

Bowman, Nicholas David, Joseph Wasserman, and Jaime Banks (2018). "Development of the video game demand scale". In: *Video Games*. Routledge, pp. 208–233.

Calderón, Alejandro and Mercedes Ruiz (2015). "A systematic literature review on serious games evaluation: An application to software project management". In: *Computers & Education* 87, pp. 396–422.

Carrión-Toro, Mayra et al. (2020). "iPlus a user-centered methodology for serious games design". In: *Applied Sciences* 10.24, p. 9007.

Casassus, Martin et al. (2019). “Time Perception and autistic spectrum condition: A systematic review”. In: *Autism Research* 12.10, pp. 1440–1462. DOI: 10.1002/aur.2170.

Chang, Chi-Cheng and Tristan Johnson (2021). “Integrating heuristics and think-aloud approach to evaluate the usability of game-based learning material”. In: *Journal of Computers in Education* 8, pp. 137–157.

Chung, Lawrence et al. (2012). *Non-functional requirements in software engineering*. Vol. 5. Springer Science & Business Media.

Condori-Fernandez, Nelly and Patricia Lago (2018). URL: https://research.vu.nl/files/68774558/Sustainability_quality_Model_version_1_0.pdf.

Condori-Fernández, Nelly et al. (2013). “An empirical approach for evaluating the usability of model-driven tools”. In: *Science of Computer Programming* 78.11. Special section on Mathematics of Program Construction (MPC 2010) and Special section on methodological development of interactive systems from Interaccion 2011, pp. 2245–2258. ISSN: 0167-6423. DOI: https://doi.org/10.1016/j.scico.2012.07.017. URL: https://www.sciencedirect.com/science/article/pii/S0167642312001451.

Connolly, Thomas M. et al. (2012). “A systematic literature review of empirical evidence on computer games and serious games”. In: *Computers Education* 59.2, pp. 661–686. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2012.03.004. URL: https://www.sciencedirect.com/science/article/pii/S0360131512000619.

Cowley, Benjamin et al. (Jan. 2014). “Learning When Serious: Psychophysiological Evaluation of a Technology-Enhanced Learning Game”. In: *under review - Journal of Educational Technology Society* 17.

De Kort, Yvonne AW, Wijnand A IJsselsteijn, and Karolien Poels (2007). “Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ)”. In: *Proceedings of PRESENCE* 195203, pp. 1–9.

Denisova, Alena et al. (2020). “Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS)”. In: *International Journal of Human-Computer Studies* 137, p. 102383.

Desmet, PMA (2000). “Emotion through expression: Designing mobile telephones with An emotional fit”. In: *Report of modeling the evaluation structure of Kansei*.

Desurvire, Heather, Martin Caplan, and Jozsef A Toth (2004). “Using heuristics to evaluate the playability of games”. In: *CHI’04 extended abstracts on Human factors in computing systems*, pp. 1509–1512.

Drachen, Anders, Magy El-Nasr, and Alessandro Canossa (Mar. 2013). “Game Analytics – The Basics”. In: *Game Analytics Maximizing the Value of Player Data*, pp. 13–40. DOI: 10.1007/978-1-4471-4769-5_2.

Duboc, Leticia et al. (2020). “Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow”. In: *Requirements Engineering* 25, pp. 469–492.

Ekman, Paul and Wallace V Friesen (1978). “Facial action coding system”. In: *Environmental Psychology & Nonverbal Behavior*.

Ekman, Paul, Wallace V Friesen, and Silvan S Tomkins (1971). “Facial affect scoring technique: A first validity study”. In.

Fu, Fong-Ling, Rong-Chang Su, and Sheng-Chin Yu (2009). “EGameFlow: A scale to measure learners’ enjoyment of e-learning games”. In: *Computers & Education* 52.1, pp. 101–112.

Gao, Lanlan, Rupert Ward, and Carlo Fabricatore (2023). “Guidelines of Serious Game Design for Promoting Reframing”. In: *Simulation & Gaming* 54.1, pp. 68–84. DOI: 10.1177/10468781221144787. eprint: https://doi.org/10.1177/10468781221144787. URL: https://doi.org/10.1177/10468781221144787.

Gris, Gabriele and Clarissa Bengtson (2021). “Assessment measures in game-based learning research: A systematic review”. In: *International Journal of Serious Games* 8.1, pp. 3–26.

HA, Mohd Noor (2012). “A review of serious game for autism children”. In: *Computer Games, Multimedia and Allied Technology (CGAT 2012)* 90.

Holloway, A and S Kurniawan (2010). “Human-centered design method for serious games: a bridge across disciplines”. In: *Proceedings from UCSC-SOE*, pp. 10–36.

IJsselsteijn, W.A., Y.A.W. de Kort, and K. Poels (2013). *The Game Experience Questionnaire*. English. Technische Universiteit Eindhoven.

IJsselsteijn, Wijnand et al. (2007). “Characterising and measuring user experiences in digital games”. In: *International conference on advances in computer entertainment technology*. Vol. 2, p. 27.

Izard, Carroll E (1984). “Emotion-cognition relationships and human development”. In: *Emotions, cognition, and behavior*, pp. 17–37.

Izard, Carroll E et al. (1993). “Stability of emotion experiences and their relations to traits of personality.” In: *Journal of personality and social psychology* 64.5, p. 847.

Johnson, Daniel, M John Gardner, and Ryan Perry (2018). “Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ)”. In: *International Journal of Human-Computer Studies* 118, pp. 38–46.

Longstreet, C Shaun and Kendra ML Cooper (2012). “Developing a meta-model for serious games in higher education”. In: *2012 IEEE 12th International Conference on Advanced Learning Technologies*. IEEE, pp. 684–685.

Machalicek, Wendy et al. (2007). “A review of interventions to reduce challenging behavior in school settings for students with autism spectrum disorders”. In: *Research in Autism Spectrum Disorders* 1.3, pp. 229–246. ISSN: 1750-9467. DOI: https://doi.org/10.

1016/j.rasd.2006.10.005. URL: https://www.sciencedirect.com/science/article/pii/S1750946706000201.

Mamani, Yonel (2021). *Deteccion de estres en tiempo real a partir de señales de voz y datos fisiologicos – repositorio.unsa.edu.pe*. http://hdl.handle.net/20.500.12773/15284. Bachelor's Thesis.

Marciano, Juvane Nunes, Leonardo Cunha de Miranda, and Erica Esteves Cunha de Miranda (2015). "Evaluating multiple aspects of educational computer games: literature review and case study". In: *International Journal of Computer Games Technology* 2014, pp. 14–14.

Maria Kellidou, Pavlina et al. (2021a). "A Review of Digital Games for Children with Autism Spectrum Disorder". In: *Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*. DSAI '20. Online, Portugal: Association for Computing Machinery, pp. 227–234. ISBN: 9781450389372. DOI: 10.1145/3439231.3439270. URL: https://doi.org/10.1145/3439231.3439270.

– (2021b). "A Review of Digital Games for Children with Autism Spectrum Disorder". In: *Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*. DSAI '20. Online, Portugal: Association for Computing Machinery, pp. 227–234. ISBN: 9781450389372. DOI: 10.1145/3439231.3439270. URL: https://doi-org.ezproxy.cc.lut.fi/10.1145/3439231.3439270.

McNair, Douglas M, Maurice Lorr, Leo F Droppleman, et al. (1971). "Manual profile of mood states". In.

Miura, Takahiro et al. (Dec. 2020). "GAME: Game As a Measurement Environment: Scheme to Evaluate Interfaces and Game Contents Based on Test Theories". In: 4.4. DOI: 10.1145/3432702. URL: https://doi.org/10.1145/3432702.

Mohamed, Hasiah and Azizah Jaafar (2010). "Development and potential analysis of heuristic evaluation for educational computer game (PHEG)". In: *5th international conference on computer sciences and convergence information technology*. IEEE, pp. 222–227.

Mohamed, Hasiah, Rohana Yusoff, and Azizah Jaafar (2012). "Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG)". In: *2012 International Conference on Information Retrieval & Knowledge Management*. IEEE, pp. 187–192.

Nacke, Lennart E (2015). "Games user research and physiological game evaluation". In: *Game user experience evaluation*, pp. 63–86.

Nowlis, Vincent (1965). "Research with the mood adjective check list." In.

Öhman, Arne and Joaquim JF Soares (1994). "" Unconscious anxiety": phobic responses to masked stimuli." In: *Journal of abnormal psychology* 103.2, p. 231.

Omar, Hasiah Mohamed, Roslina Ibrahim, and Azizah Jaafar (2011). "Methodology to evaluate interface of educational computer game". In: *2011 International Conference on Pattern Analysis and Intelligence Robotics*. Vol. 2. IEEE, pp. 228–232.

Omar, HM and A Jaafar (2008). "Playability Heuristics Evaluation (PHE) approach for Malaysian educational games". In: *2008 International Symposium on Information Technology*. Vol. 3. IEEE, pp. 1–7.

Overskeid, Geir (2016). "Power and Autistic Traits". In: *Frontiers in Psychology* 7. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2016.01290. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01290.

Pagulayan, Randy J et al. (2007). "User-centered design in games". In: *The human-computer interaction handbook*. CRC Press, pp. 767–786.

Pandeliev, Velian T. and Ronald M. Baecker (2010). "A Framework for the Online Evaluation of Serious Games". In: *Proceedings of the International Academic Conference on the Future of Game Design and Technology*. Futureplay '10. Vancouver, British Columbia, Canada: Association for Computing Machinery, pp. 239–242. ISBN: 9781450302357. DOI: 10.1145/1920778.1920816. URL: https://doi.org/10.1145/1920778.1920816.

Petri, Giani and Christiane Gresse von Wangenheim (2016). "How to evaluate educational games: a systematic". In: *Journal of Universal Computer Science* 22.7, pp. 992–1021.

Petri, Giani, Christiane Gresse von Wangenheim, and Adriano Ferreti Borgatto (2019). *MEEGA+, Systematic Model to Evaluate Educational Games.*

Razali, Nurlieda Ellyanna Munirrah et al. (2022). "Identifying and validating game design elements in serious game guideline for climate change". In: *Heliyon* 8.1, e08773. ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2022.e08773. URL: https://www.sciencedirect.com/science/article/pii/S2405844022000615.

Ritterfeld, Ute, Michael Cody, and Peter Vorderer (2009). *Serious games: Mechanisms and effects*. Routledge.

Ryan, Richard M, C Scott Rigby, and Andrew Przybylski (2006). "The motivational pull of video games: A self-determination theory approach". In: *Motivation and emotion* 30, pp. 344–360.

Sucksmith, Ed, Ilona Roth, and Rosa Anna Hoekstra (2011). "Autistic traits below the clinical threshold: re-examining the broader autism phenotype in the 21st century". In: *Neuropsychology review* 21, pp. 360–389.

Suits, Bernard (1967). "What is a Game?" In: *Philosophy of Science* 34.2, pp. 148–156. DOI: 10.1086/288138.

Suni Lopez, Franci, Nelly Condori-Fernandez, and Alejandro Catala (2019). "Towards real-time automatic stress detection for office workplaces". In: *Information Management and*

Big Data: 5th International Conference, SIMBig 2018, Lima, Peru, September 3–5, 2018, Proceedings 5. Springer, pp. 273–288.

Sweetser, Penelope and Peta Wyeth (2005). "GameFlow: a model for evaluating player enjoyment in games". In: *Computers in Entertainment (CIE)* 3.3, pp. 3–3.

Szilas, Nicolas (2022). "Serious Game Design in Practice: Lessons Learned from a Corpus of Games Developed in an Academic Context". In: *Proceedings of the 17th International Conference on the Foundations of Digital Games*. FDG '22. Athens, Greece: Association for Computing Machinery. ISBN: 9781450397957. DOI: 10.1145/3555858.3563283. URL: https://doi.org/10.1145/3555858.3563283.

Tsikinas, Stavros and Stelios Xinogalos (2018). "Studying the effects of computer serious games on people with intellectual disabilities or autism spectrum disorder: A systematic literature review". In: *Journal of Computer Assisted Learning* 35.1, pp. 61–73. DOI: 10.1111/jcal.12311.

– (2019). "Design guidelines for serious games targeted to people with autism". In: *Smart Education and e-Learning 2019*. Springer, pp. 489–499.

Urbanek, Michael and Florian Güldenpfennig (2019). "Celebrating 20 years of computer-based audio gaming". In: *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, pp. 90–97.

Waes, L. van (2000). "Thinking aloud as a method for testing the usability of Websites: the influence of task variation on the evaluation of hypertext". In: *IEEE Transactions on Professional Communication* 43.3, pp. 279–291. DOI: 10.1109/47.867944.

Watson, David and Lee Anna Clark (1994). "The PANAS-X: Manual for the positive and negative affect schedule-expanded form". In.

White, Susan W. et al. (2009). "Anxiety in children and adolescents with autism spectrum disorders". In: *Clinical Psychology Review* 29.3, pp. 216–229. ISSN: 0272-7358. DOI: https://doi.org/10.1016/j.cpr.2009.01.003. URL: https://www.sciencedirect.com/science/article/pii/S027273580900004X.

Wieringa, Roel J (2014). *Design science methodology for information systems and software engineering*. Springer.

Wohlin, Claes (2014). "Guidelines for snowballing in systematic literature studies and a replication in software engineering". In: *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pp. 1–10.

Zhonggen, Yu (2019). "A meta-analysis of use of serious games in education over a decade". In: *International Journal of Computer Games Technology* 2019, pp. 1–8. DOI: 10.1155/2019/4797032.

Zuckerman, Marvin, Benard Lubin, and Christine M Rinck (1983). "Construction of new scales for the multiple affect adjective check list". In: *Journal of Behavioral Assessment* 5, pp. 119–129.

A APPENDIX: PROPOSED SERIOUS GAMES (SG) LOGS SCRIPT

```
using System;
using System.Collections;
using System.Collections.Generic;
using System.IO;
using UnityEngine.SceneManagement;
using UnityEngine;
using UnityEngine.EventSystems;
using UnityEngine.UI;

public class MakeJsonLogs{
    [System.Serializable]
    public class Logs
    {
        public string logTime = System.DateTime.Now.ToString("yyyy-MM-dd HH:mm:ss");
        public string scene;
        public string asset;
        public string events;
        //public string activity;
        public string task;
    }

    [System.Serializable]
    public class LogsList
    {
        public Logs[] logs;
    }

    public LogsList myLogsList = new LogsList();
    public Logs myLogs = new Logs();
    public void MakeFile()
    {
        string logsJson = JsonUtility.ToJson(myLogsList, true);
        File.WriteAllText(Application.dataPath + "/PlayerGameLogs.json", logsJson);
    }

    public void AddRecord(string events, string task)
    {
        //read json from text file
```

```
        string str = File.ReadAllText(Application.dataPath + "/PlayerGameLogs.json");
        myLogsList = JsonUtility.FromJson<LogsList>(str);
        //collect data from input fields
        Logs myPlayerLogs = new Logs();
        myPlayerLogs.logTime = System.DateTime.Now.ToString("yyyy-MM-dd HH:mm:ss");
        myPlayerLogs.scene = EventSystem.current.currentSelectedGameObject.tag;
        myPlayerLogs.asset = EventSystem.current.currentSelectedGameObject.name;
        //myPlayerLogs.events = EventSystem.current.currentSelectedGameObject.name + " is clicked";
        myPlayerLogs.events = events;
        myPlayerLogs.task = task;
        //make temp list to hold old data
        LogsList myLogsListTemp = new LogsList();
        myLogsListTemp.logs = myLogsList.logs;
        //make original longer
        myLogsList.logs = new Logs [myLogsListTemp.logs.Length + 1];
        //copy old data across
        for (int i = 0; i < myLogsListTemp.logs.Length; i++)
        {
            myLogsList.logs[i] = myLogsListTemp.logs[i];
        }
        //add new data
        myLogsList.logs[myLogsListTemp.logs.Length] = myPlayerLogs;
        //write data
        outputJSON();
    }

    public void outputJSON()
    {
        string strOutput = JsonUtility.ToJson(myLogsList);
        File.WriteAllText(Application.dataPath + "/PlayerGameLogs.json", strOutput);
    }
}
```

Listing A.1: MakeJsonLogs

```
using System;
using System.Collections;
using System.Collections.Generic;
using System.IO;
using UnityEngine.SceneManagement;
using UnityEngine;
```

```
using UnityEngine.EventSystems;
using UnityEngine.UI;

public class menuutama : MonoBehaviour
{
  public Button warna;
  public Button kenalbangun;
  public Button ayobersikap;
  public Button keluar;
  public Button kembali;
  public Button bantuan;
  public Canvas panelbantuan;
  public MakeJsonLogs mj = new MakeJsonLogs();
  void Start()
  {
    warna = warna.GetComponent<Button>();
    kenalbangun = kenalbangun.GetComponent<Button>();
    ayobersikap = ayobersikap.GetComponent<Button>();
    keluar = keluar.GetComponent<Button>();
    mj.AddRecord("Player is in main menu", "");
  }

  public void exit()
  {
    mj.AddRecord("Player goes to closing scene", "");
    SceneManager.LoadScene("end");
  }

  public void color()
  {
    mj.AddRecord("Player goes to coloring scene", "");
    SceneManager.LoadScene("warna");
  }

  public void shape()
  {
    mj.AddRecord("Player goes to study shape menu", "");
    SceneManager.LoadScene("kenal bangun");
  }

  public void behave()
  {
    mj.AddRecord("Player goes to acts scene", "");
    SceneManager.LoadScene("bersikap");
  }}
```

Listing A.2: Example of calling MakeJsonLogs from main menu class

```
using UnityEngine;
using UnityEngine.Audio;
using System.Collections;

public class dragdrop : MonoBehaviour {
  public MakeJsonLogs mj;
  Vector3 initialPosition1;
  public GameObject soal, Jawab;
  public AudioSource benar;
  public AudioSource salah;
  public Canvas myCanvas;
  private int x;
  Camera camera;
  Vector3 screenPos;
  Vector2 pos;

  void Start () {
    initialPosition1 = soal.transform.position;
  }
  // Update is called once per frame
  void Update () {
    screenPos = camera.WorldToScreenPoint(Jawab.transform.position);
  }

  public void Drag()
  {
    mj = new MakeJsonLogs();
    mj.AddRecord("Player is dragging asset", "");
    RectTransformUtility.ScreenPointToLocalPointInRectangle(myCanvas.
   transform as RectTransform, Input.mousePosition, myCanvas.worldCamera,
    out pos);
    soal.transform.position = myCanvas.transform.TransformPoint(pos);
  }

  public void Drop()
  {
    mj = new MakeJsonLogs();
    float distance1 = Vector3.Distance (soal.transform.position, Jawab.
   transform.position);

    if (distance1 < 10) {
      soal.transform.position = Jawab.transform.position;
      benar.Play ();
      salah.Stop ();
      mj.AddRecord("Player is dropping asset", "success");
    } else {
```

```
44          soal.transform.position = initialPosition1;
45          benar.Stop ();
46          salah.Play ();
47          mj.AddRecord("Player is dropping asset", "fail");
48      }
49    }
50  }
```

Listing A.3: Example of calling MakeJsonLogs for scene with task completion

B APPENDIX: LIST OF TASKS

Following is the full list of tasks and activities for of GVT usabilities:

- Desktop app

  T1 Configure GVT

  - You are a serious game developer, wants to validate your game using GVT
  - You want to entry the player's name that ready to test your game using GVT
  - You want to locate your game using GVT

  A1.1 Modify GVT entries

  - You wrongly input the player's name and game path, therefore, you want to reset the GVT entry columns.

  A1.2 Starting GVT

  - You want to start GVT

  A1.3 Stopping GVT

  - You want to start GVT

- Web App

  T2 Evaluate Single Analysis page

  A2.1 Monitoring single player's records

  - You have only one player ready to test and validate your game
  - Use GVT web app to monitor the player's records

  A2.2 Download Monitoring Data table

  - You want to download Monitoring Data table for further documentation

  A2.3 Analyse single player's records

  - You want to see the more detailed analysis with the time spent value and time accumulation from the game testing records by GVT

  A2.4 Filter and sort Single Analysis Table

  - You want to filter and sort the Single Analysis table to show you your desire table
  - You want to sort the table based on the time in ascending order
  - You want to sort the table based on the time in descending order
  - You want to sort the table based on the scene's name in ascending order
  - You want to sort the table based on the scene's name in descending order
  - You want to sort the table based on the asset's name in ascending order

- You want to sort the table based on the asset's name in descending order
- You want to sort the table based on the events' name in ascending order
- You want to sort the table based on the events' name in descending order
- You want to sort the table based on the time spent in ascending order
- You want to sort the table based on the time spent in descending order
- You want to sort the table based on the time accumulation in ascending order
- You want to sort the table based on the time accumulation in descending order
- You want to sort the table based on the stress level in ascending order
- You want to sort the table based on the stress level in descending order
- You want to filter the table to only show the records which the task completion is "success"
- You want to filter the table to only show the records which the task completion is "fail"

A2.5 Download Single Analysis table

- You want to download a table with the Time Spent and Time Accumulation values

A2.6 Grouping the records based on scene

- You want to see all the records based on their scenes (not the individual records), so that you can analyse the total time spent in each particular scene and the average stress level in each scene

A2.7 Download records based on scene

- You want to download all the grouping records by their scene, for further documentation

A2.8 Show the analysis as a Graph

- You want to see the analysis result as a graph
- You want to see a line graph that shows the relation between Stress Level and Scene
- You want to see a bar graph that shows the relation between Stress Level and Scene
- You want to see a line graph that shows the relation between Total Time and Scene
- You want to see a bar graph that shows the relation between Total Time and Scene

- You want to see a line graph that shows the relation between Stress Level and Total Time
- You want to see a bar graph that shows the relation between Stress Level and Total Time

A2.9 Read the Glossary for single analysis

- You want to know how GVT done the analysis from the records and what formula GVT used to calculate and produce the result

T3 Evaluate Multi Analysis page

A3.1 Do Multi Analysis

- After done testing your game with only one player, you now have multiple players ready to testing your game
- You want to see the final result of your game feasibility from all the records you gathered from many players.

A3.2 Show the result as a table

- You want to see the result in more detail as a table

A3.3 Download Result Detail table

- You want to download the detailed result table as xls file for further documentation

A3.4 Show the Result Detail table as a graph

- You want to show the Result Detail table as a graph
- You want to see a line graph that shows the relation between Avg Stress Level and Scene
- You want to see a bar graph that shows the relation between Avg Stress Level and Scene
- You want to see a line graph that shows the relation between Avg Total Time and Scene
- You want to see a bar graph that shows the relation between Avg Total Time and Scene
- You want to see a line graph that shows the relation between Avg Stress Level and Total Time
- You want to see a bar graph that shows the relation between Avg Stress Level and Total Time

A3.5 Compare the multi user's records in a chart

- You want to see a multi-line graph that shows the relation between Avg Stress Level and Scene to compare multi user's records

- You want to see a multi-bar graph that shows the relation between Avg Stress Level and Scene to compare multi user's records
- You want to see a multi-line graph that shows the relation between Avg Total Time and Scene to compare multi user's records
- You want to see a multi-bar graph that shows the relation between Avg Total Time and Scene to compare multi user's records

A3.6 Read the Glossary for multi analysis

- You want to see the formula to calculate the multi analysis result

C APPENDIX: QUESTIONNAIRE - LIST OF QUESTIONS

Two questions and one entry suggestion were presented in the GVT desktop app. The questions are as follows:

- Is the configuration of GVT desktop app manageable?
- Are you satisfied with the GVT desktop app?
- Please give your further comment about GVT desktop app

The list of questions for evaluating GVT web app are as follows:

- Home
  - Is the Home manageable?
  - Are you satisfied with the Home page?
  - Please give your further comment about this page
- Single Analysis
  - Is the Single Analysis page manageable?
  - Are you satisfied with the Single Analysis page?
  - Is Monitoring Data table easy to understand?
  - Are you satisfied with the Monitoring Data table?
  - Is Single Analysis table easy to understand?
  - Is Single Analysis table manageable?
  - Are you satisfied with the Single Analysis table?
  - Is Analysis Based on Scene table easy to understand?
  - Are you satisfied with the Analysis Based on Scene table?
  - Is Graph Representative easy to understand?
  - Is Graph Representative manageable?
  - Are you satisfied with Graph Representative?
  - Is the Glossary easy to understand?
  - Are you satisfied with the Glossary
  - Please give your further comment about this page
- Multi Analysis
  - Is the Multi Analysis page manageable?
  - Are you satisfied with the Multi Analysis page?
  - Is the Assessment Result easy to understand?
  - Are you satisfied with the Assessment Result?

- Is the Result Detail table easy to understand?
- Are you satisfied with the Result Detail table?
- Is the Result Graph Representation easy to understand?
- Is the Result Graph Representation manageable?
- Are you satisfied with the Result Graph representation?
- Is the Detailed Graph Representation - Multi Players easy to understand?
- Is the Detailed Graph Representation - Multi Players manageable?
- Are you satisfied with the Detailed Graph Representation - Multi Players?
- Is the Glossary easy to understand?
- Are you satisfied with the Glossary?
- Please give your further comment about this page

• About
  - Is About page easy to understand?
  - Are you satisfied with the About page?
  - Please give your further comment about this page

D APPENDIX: GVT SOURCE CODE

The full source code of the GVT can be found in the github repository: GVT