Using learning analytics to understand esports students

Daniel Ostnes

30.05.2021

Supervisor: Barbara Wasson
Co-supervisor: Angelica B. Ortiz de Gortari

University of Bergen
Faculty of Social Sciences
Department of Information Science and Media Studies
Centre for the Science of Learning & Technology
Master Thesis

"I see now that the circumstances of one's birth are irrelevant. It is what you do with the gift of life that determines who you are" **- Mewtwo, 1999**

Abstract

Electronic sports (esports) has advanced to become a media giant and an arena for competitions and career development. Due to this growth, more focus has been given to esports research, implementation of esports throughout the world, and development of esports curriculum. Introducing esports into schools has created huge opportunities for deeper analysis of esport and learning data to provide insight into the learning processes. By applying learning analytics methods, this research analyzes data that originate from students (N=149) in Swedish high schools. The data was divided between activity data and performance data. The analysis is guided by the learning theory concept self-regulation to analyze differences between user groups. Through exploratory analysis, multiple user groups were identified and then compared in their trends and results to measure the impact of self-regulated learning concepts. Furthermore, the student data was used in the design of a mid-fidelity prototype for a student-facing dashboard to provide feedback and recommendations.

Findings reveal that concepts of self-regulated learning have a positive impact in terms of higher curriculum interaction, and also higher performance results in game matches. While the research finds that focus on features promoting self-regulated learning concepts is important, it is challenging to generalize the findings to recommend actions such as suggested session lengths. Future work should include a larger population sample and focus on the implementation of a student-facing dashboard tool to test its reception and usage.

Acknowledgments

First and foremost, I would like to extend my sincerest gratitude to my supervisor Barbara Wasson and co-supervisor Angelica B. Ortiz de Gortari. Your expertise, help, and constructive criticism has been invaluable. A great many thanks to Jonathan Bäckström from Learn2Esport for your technical support and expertise.

Secondly, my friends and family have been my structural bedrock when motivation and progression seemed afar. A special thanks to Benjamin Andersen for our newfound contact and online sessions during the pandemic, and Sunniva Stolt-Nielsen for a new friendship that I greatly look forward to.

And last but not least, I would also like to thank my better half, Camilla, for believing in me and supporting me in all my endeavors. Your continuous support and encouragement has been invaluable.

Daniel Ostnes

30.05.2021

Contents

List of Tables

List of Figures

Abbreviations

esports: Electronic sports

IESF: International Esports Federation

LA: Learning analytics

LAD: Learning analytic dashboards

LMS: Learning management systems

SRL: Self-regulated learning

OLM: Open learner models

CS:GO: Counter Strike: Global Offensive

CT: Counter terrorist

T: Terrorist

FPS: First person shooter

LoL: League of Legends

MOBA: Multiplayer online battle arena

$\mathbf{N_{PL}}$: Profile linking

$\mathbf{N_{NPL}}$: Non-profile linking

$\mathbf{N_{JE}}$: Journal entry

$\mathbf{N_{NJE}}$: Non-journal entry

1 Introduction

Electronic sports (esports) has evolved from a recreational activity to a media giant and an arena for competitions and career development. In 2020 it was reported that more than 495 million people frequently watched esports or played it themselves (Newzoo, 2020). Due to the growth of esport enthusiasm there has been an increased interest in esports research, the development of esports curriculum, and a discussion on *if esports can be considered a sport*. The introduction of data from esports in schools has created a huge opportunity for deep analytics and personalization of the learning processes. Normally for learning platforms there are many variables and conclusions drawn from data, such as time spent on task, the number of clicks, and forum activity. Data from esports activities has the unique opportunity to provide more extensive insight into the effect of time spent on activities such as curriculum interaction, but also time spent while playing. Combining esports data with exam results, self-reported data, and other data can supply students with rich feedback on their performance and trends, and recommend actions for further progression. Just as data is aiding other sports such as motor racing and football clubs to improve their performance, esports is well-conditioned to follow in the same steps. Winning in esports is not just determined by skill levels or dedication, but employing strategies and analysis of past performance is crucial to solve the secrets to success (Wooden, n.d.).

1.1 Motivation

Being born in the 90s, I have grown up alongside a massive technological growth, and I quickly found my passion for video games from a very young age. Most societies have been through a *"video games are bad"* phase motivated by a fear that kids might grow up to be violent and develop squared eyes. Despite this, it seems that video games are becoming accepted more than just a hobby; but as a career choice or as a platform for learning. Being presented with the opportunity to research the usefulness of esports data seemed like a choice too good to pass. Combining this with my enthusiasm for data analysis and as an advocate for educational development, this research project has been engaging and valuable.

1.2 Research Questions

Working with learning analytics within an esports environment produces many opportunities and challenges. It is a data rich field, and as esports is gaining popularity and seeing implementation in school settings, it is important to approach the data with specific goals in mind. First, this research covers the use of learning theories and learning concepts used by students who enroll in an esports course at high schools in Sweden. Approaching the data with a pedagogical view can strengthen the use of esports data, and lead to insights to further develop curriculum and technology based on learning and progression. Second, the data can be analyzed and used in student-facing dashboards for students to receive feedback, instructions, and recommendations. This may be challenging, both in terms of ethics, but also *how* data should be presented.

Last, for esports and video games in general, it may be hard to identify how performance is related to session lengths. Analyzing data related to activity and performance may produce recommendations about how much time should be spent on training (e.g., X is too many, Y is too few, Z is balanced).

Based on the research goal *"To investigate how learning analytics can be used to provide insights into various aspects of esports education"*, the following research questions are asked:

RQ1: How are self-regulated concepts related to esports performance and activity?

RQ2: What can learning analytics tell us about esports students?

RQ3: How are session lengths tied to performance?

1.3 Contribution

This Master research looks into learning analytics methods and its utilization on rich esports data that originate from Swedish high schools. Studying learning in technology enhanced environments can be challenging, and more often than not researchers and stakeholders may be blinded by data availability instead of keeping learning theories and pedagogy in mind. Instead of falling for the temptation of data availability, learning analytics should focus on the interests of the learner and to reveal new insights to support learning (Drachsler & Greller, 2012). This project explores the depths of student data that stems from an educational esports environment and how the usage of learning theory concepts can help distinguish students and their performance levels. Through the development of a mid-fidelity prototype, visualizations are provided as to how esports data may be combined and used in a student-facing dashboard to provide feedback and recommendations. The thesis contributes with the following:

- A literature review of esports, learning theories for digital environments, combination of multiple datasets, and student-facing dashboards.
- Development of proxy variables to represent aspects of student activity otherwise not available in the data
- Creation of a codebook for qualitative analysis of unstructured data input
- Analysis of esports student data through user grouping, descriptive statistics, trend analysis, and thematic analysis
- Development of a mid-fidelity prototype of a student-facing dashboard

1.4 Thesis Contents

This thesis is structured into 7 chapters and this is the outline.

Chapter 2: Background presents a literature review of related fields and disciplines.

Chapter 3: Methodology gives an overview of the methodology and methods that were used.

Chapter 4: The Dataset describes the dataset, data properties, and data processing.

Chapter 5: Results presents the analysis results.

Chapter 6: Visualizations presents visualizations for a student facing dashboard application as a mid-fidelity prototype.

Chapter 7: Discussion is a discussion of the research approach, results, and limitations.

Chapter 8: Conclusion presents the conclusion of the research and future work.

2 Background

This chapter presents background information related to this research to place fields and definitions in context. An overview of query design and query limitations is presented. Reviewing disciplines related to esports included learning theories, learning management systems, and methods to analyze and work with academic data. As this research focus on an educational environment it is important to include academic approaches (learning theories and self-regulated learning), how to handle the student data (learning management systems, learning analytics, open learner models). With esports being in a digital environment it is essential to attend to its surrounding technology and concerns while maintaining an academic setting. A review of the included video games in this research is also given.

2.1 Query Design

For this research, multiple databases and two journals have been queried, namely: Google Scholar[1], Web of Science[2], LAK: Learning Analytics & Knowledge Conferences[3], Proceedings from the International Conference on Quantitative Ethnography[4].

The scope of this project includes preliminary work from a project proposal, and some information and articles have been found where the search query is not available (as far as the literature review and background theory is concerned). Table 1 has an overview of keywords that were used in the literature search.

Table 1: Query Keyword Overview

Theme	Keywords
Esport	Esports, Esports in school, sports, sports science,
School	school, education, curriculum, learning, learn, learner, student
Data	self-report, self report, self reporting, unstructured, journals, diaries, diary, models, data, visualization, statistics, text
Theory	learning analytics, analytics, analysis, learner models, learning management system, esports management system, self-regulated, self regulated, self regulation, open learner models, thematic analysis

Google Scholar mainly produces results that are English, and these were automatically sorted by citation count and general relevance. It is not abnormal to retrieve many results on Google Scholar as it includes articles, reports, books, direct citations, and much more. Because of its focus on computer science and social sciences, Web of Science was a relevant database to query. The results are not as far-reaching as Google Scholar, but Web of Science calculates an impact factor, and also produces very accurate results if the queries are well-defined.

[1] https://scholar.google.com
[2] https://apps.webofknowledge.com/
[3] https://dl.acm.org/conference/lak
[4] https://qesoc.org

The Learning Analytics & Knowledge Conference is a yearly conference with proceedings and publications that focus on Learning Analytics. Since its first proceedings in 2011, there have been annual conferences with many publications and significant researchers have taken part in pushing the state of the art of Learning Analytics through this fora. The International Conference on Quantitative Ethnography is a new annual conference since 2019 that focus on research of human thought, behavior, and interaction, and one of its fields are within education. All publications and proceedings since for the conferences have been screened and included if found consistent according to the research aim.

The literature search was limited to results written in English or Norwegian. No specific year limit was used due to full inclusion of well-established fields and seminal work in mature disciplines. Voiding a specific year limit assures the inclusion of (often) older and fundamental aspects on the one hand, and on the other hand includes new work that may push the state of the art. Logical operators (i.e., AND, OR) were used to combine the keywords to structure more advanced queries to limit results. To query exact matches, words or phrases were put into quotes (e.g., "esports"). Search results were considered useful if they met criteria to strengthen the research inquiry:

- Esports in school, or esports curriculum
- Learning analytics, student data, learner models, and learning management systems
- Learning theories in digital learning environments
- Student facing dashboards and how to present data

In the preliminary phase of the research, keywords were queried in isolation to produce a high number of results to both assess the disciplines and to gain knowledge. Due to this, the total number of hits for the literature search is unavailable. Through a relevancy process (Figure 1), a total of 132 references were considered. Further screening (full reading) resulted in 69 core references for the project. There is a total of 87 bibliography entries, but not all entries originate from the literature search. 69 out of the 87 references (79.31%) are articles, conference proceedings, or books. The remaining 18 (20.69%) are websites, master theses, reports, or journals.

2.2 Query Limitations

This research is concerned with Learning Analytics and the emergence of esports in education. Learning Analytics is a new but quickly evolving field that have some common research disciplines such as educational data mining and open learner models. Thus, through a grounded approach to the inclusion of topics and their connectivity, there may be limitations to exclusion of other relevant fields. As with esports, it is a new phenomenon to consider its utilization in an educational setting. There is limited research produced for esports in education, thus the research has included esports in combination with self-regulated learning, self-reporting of performance, esports as a sport, trace data, visualization of data, learner models and techniques to open learner models, and methods of learning analytics data handling.

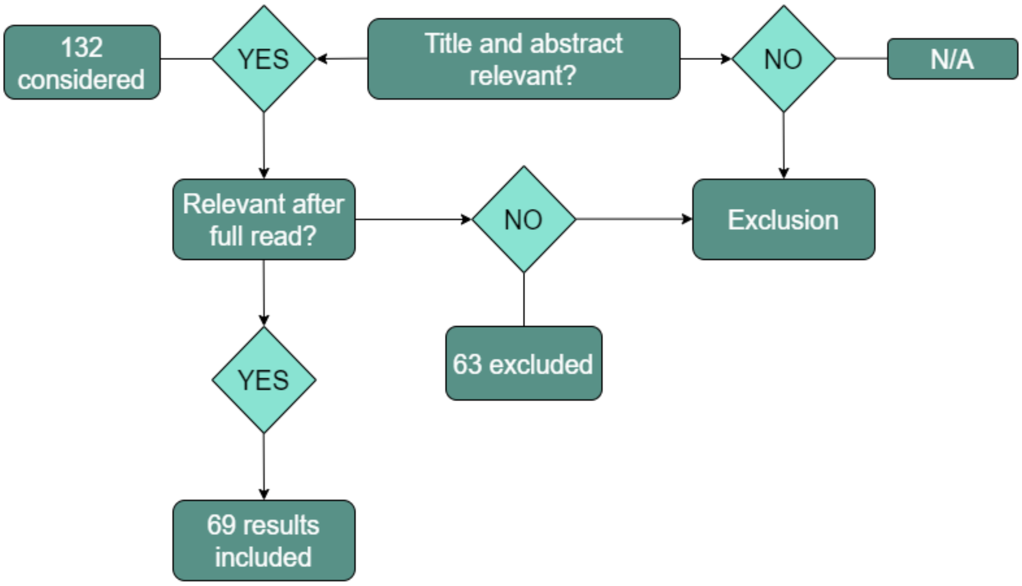

Figure 1: Literature Search Relevancy Process

2.3 Esports

Electronics sports (henceforth denoted esports[5]) has taken video games and the gaming industry from a recreational activity to an enormous competitive arena and media giant. Esports refers to "situations where computer games are played competitively" (Schubert, Drachen, & Mahlmann, 2016). Esports is defined as:

> *"the casual or organized competitive activity of playing specific video games that provide professional and/or personal development to the player. This practice is facilitated by electronic systems, either computers, consoles, tablets, or mobile phones, on which teams and individual players practice and compete online and/or in local-area-network-tournaments at the professional or amateur level. The games are established by ranking systems and competitions and are regulated by official leagues. This structure provides players a sense of being part of a community and facilitates mastering expertise in fine-motor coordination and perceptual-cognitive skills, particularly but not exclusively, at higher levels of performance"* (Pedraza-Ramirez, Musculus, Raab, & Laborde, 2020).

Following this definition, it is worth noting the differences between a video game and esports video game. If a game does not have official leagues that regulate competitions, or an existing ranking system, they may per definition not be categorized as an esports game. Not every video game is an esports game, but every esports game is a video game (Pedraza-Ramirez et al., 2020). In recent years, esports has evolved to be one of the most popular media forms (Hamari & Sjöblom, 2017), and in 2020 it was reported that more than 495 million people occasionally watched broadcasts or partook in esports activities (Newzoo, 2020).

[5]The spelling of esports come in many forms (e.g., Esports, eSports, e-sports, eSports) - but this thesis will use the spelling esports, with regular capitalization after a period. The Associated Press added 'esports' as an entry to their style book where they previously defined 'email' - denouncing the use of capitalized 'E' or hyphenation. The use of alternate forms are only if part of a formal name, like an organization (ESPN, 2017).

The 2020 Esports Market Report of Newzoo also expects an audience growth to 645 million by the end of 2023 with a compound annual growth rate of 10.4% from 2018 to 2023 (Newzoo, 2020). Esports awareness and esports enthusiasts both saw an increase of 11.27% and 10.79%, respectively from 2019 to 2020, while esports revenues stayed about the same for both years. The reports show a decrease in total hours watched live from 1265.8 million hours in 2019 to 1209.6 million hours in 2020. This can be explained by the difference between esports-aware people and esports enthusiasts, and the interchangeable nature of being enthusiastically involved in any given topic. Total hours watched non-esports, but gaming related, saw an increase from 6607.3 million hours in 2019 to 7833.2 million hours in 2020 (Newzoo, 2019, 2020).

The growth in gaming related content will naturally bring in more awareness and enthusiasts for esports, and the growth of esports can be explained by its accessibility online and its use of broadcasting technologies. Following such advancements, there has been increased interest in esports research within fields such as sports psychology (Bányai, Griffiths, Király, & Demetrovics, 2019; Pedraza-Ramirez et al., 2020), sports analytics (Nagyl, 2016; Schubert et al., 2016; Wagner, 2006) and information technology (Hamari & Sjöblom, 2017).

2.3.1 Esports Analytics

Emerging as a multidisciplinary field of research, esports analytics investigates behavior related to esports performance and the stakeholder levels involved (e.g., personal, teams, institutions, countries). Esports analytics is based on research methods and scientific approaches developed in sports science, game analytics, learning analytics, psychology, social sciences, and other areas. Schubert et al. (2016) describe esports analytics as

> *"the process of using esports related data, primarily behavioral telemetry but also other sources, to find meaningful patterns and trends in said data, and the communication of these patterns using visualization techniques to assist with decision-making processes".*

2.3.2 Esports As A Sport

An ongoing dilemma surrounding esports is its categorization within sports. We must first define sports to further debate this. Wagner (2006) embraces the definition of sports by Tiedemann (2004) and presents a broad definition that does not alienate esports:

> *"Sport is a cultural field of activity in which people voluntarily engage with other people with the conscious intention to develop and train abilities of cultural importance and to compare themselves with these other people in these abilities according to generally accepted rules and without deliberately harming anybody".*

The notion of *what specific abilities are considered in a sport* and its cultural importance is debatable, seeing as the disagreement of esports' place in the realm of sports often emerge with new scientific publications in the field of esports.

To further add to this debate, the use of specific activities such as keyboard usage, mouse usage, hand-eye coordination, communication, decision-making, and fine motor skills (Jenny, Manning, Keiper, & Olrich, 2017) are abilities that can be trained just as any other ability that is defined in a traditionally accepted sport. Since 2016, there has been international efforts to categorize esports as a real sport through the International Esports Federation (IESF[6]). As of May 2021, IESF comprises 104 nations that consider esports a real sport, and they want to

> *"see a world where esports is accepted as a real sport and that esport athletes can compete on the same level and with the same support as athletes from traditional sports"* (IESF, n.d.).

Jenny et al. (2017) conclude that esports does have many similarities, but lacks certain aspects such as physicality and institutionalization to be categorized as a sport for some people. As closing remarks, by combining the definition of sports as presented by Wagner (2006) with the efforts of IESF it is reasonable to assume that the efforts of esports development within the sports world is a joint effort that continues to gain popularity.

2.3.3 Esports In Education

Game-based learning is reported as a valuable tool for learning (Denden, Tlili, Essalmi, & Jemni, 2018; Freire et al., 2016). Students may be pushed by intrinsic motivations by trying to get better and encourage other players to do the same. By acknowledging esports as a sport, it is possible to combine intrinsic rewards with extrinsic rewards such as prizes, payments, salaries and contracts. This may prove beneficial to discipline students in other academic performances seeing as digital games are good tools to learn language, mathematics, computer architecture, digital literacy, teamwork, communication and much more (Denden et al., 2018). These abilities align well with the thoroughly discussed topic of 21st century skills, and thus strengthens the use case for esports in education. Skills related to 21st century skills include: (1) adaptability, (2) construction of knowledge, (3) managing and understanding information, (4) critical thinking, (5) teamwork and (6) communication (Voogt & Roblin, 2010). Elias (2011) describes the development of education and knowledge creation, stating that traditional performance measures (e.g., grades) are ineffective methods to portray performance potential. The efforts of Voogt and Roblin (2010) align well with the research on *learning analytics* (see Section 2.4) by Elias (2011) on how to measure performance, stating that newer methods are more effective at estimating performance.

Lee et al. (2020) designed the first high school curriculum with esports as its focus. Through a design process, the authors integrated science, technology, engineering, mathematics, English language arts, career-technical education, and social-emotional learning as components of the curriculum. The approach was to design a curriculum based on *integration through a motivating context* (the authors identify a "crisis of literacy" and experiment with esports in academia as a solution) and esports is popular among young people and students and provides situations where reading is necessary (Lee et al., 2020). The curriculum design included students, teachers, administration, and researchers, ensuring that all stakeholders were included in the process to maintain effective communication. Development challenges were identified, see Table 2.

[6] https://ie-sf.org

Table 2: Esports Curriculum Challenges by Lee et al. (2020)

Challenge	**Definition**
Doubt	Stakeholder doubt. Specifically parents and administrators
Engagement	Student and parent engagement is important to successfully implement esports, both at a local level and a governing body
Standards	Maintaining a connection between esports and educational standards (e.g., STEM skills)
Knowledge	Ensuring proper knowledge about esports and esports related topics, specifically for teachers and coaches

Wagner (2006) discussed esports and how it conforms to academic settings. In his article, Wagner states that esports can be differentiated from sports, thus avoiding the discussion of how (or if) esports can be described as a sport and instead focus on developing esports as a discipline. However, by approaching esports as a sport, researchers can acquire approaches and methods from traditional sports science and utilize them as a basis for esports science (Wagner, 2006).

2.3.4 Cognitive Performance In Esports

Players must assess relevant information in relation to their goal (e.g., winning a round, winning a match). By processing real-time information, players must constantly commit to decisions using their memory capacity and try to determine the best outcome and act upon that. This is evidence that cognitive flexibility is abundant in esports, and new information must constantly be absorbed to create a winning solution. Further, seeing as esports games are prone to changes (e.g., visual changes, map changes, weapons changes, meta-game changes), cognitive performance is essential to master such changes. Conducting research on esports requires researchers and sports psychologists to understand the need for cognitive performance and motor skills. To achieve peak performance, high cognitive performance and higher order functions are required (Pedraza-Ramirez et al., 2020).

Students with high achievement motivation is shown to prefer competitive activities given that feedback is presented properly, and previous research also indicate that young students with traits of high achievement motivation are expected to learn competencies that align with 21st century skills during competitive video game environments (i.e., esports participation) (Fromme, 2003; Vansteenkiste & Deci, 2003; Wagner, 2006).

Pedraza-Ramirez et al. (2020) carried out a systematic review of esports psychology with an approach developed from sports psychology. The authors investigated game performance and cognitive performance in esports, and conclude that esports research can gain a lot when treated as part of sports science. As research in esports is in an introductory development phase, Pedraza-Ramirez and his co-authors want to encourage further esports performance research. A key challenge to explore is identification of performance indicators and what data can reliably provide a better understanding of esports performance (Pedraza-Ramirez et al., 2020).

2.3.5 Game Performance In Esports

There are multiple factors that define game performance in esports apart from time on task, such as pressure (e.g., prizes, time, audience) and practice (e.g., playing alone, playing with a team, boot camps). These characteristics affect a player's game performance and this requires methods from sports psychology and educational psychology to help the development. A common conception is that playing for as many hours as possible will yield good results, but previous research shows that performance decline over time (Pedraza-Ramirez et al., 2020; Sapienza, Zeng, Bessi, Lerman, & Ferrara, 2018). However, other research on video games performance does indicate that the number of matches played are the strongest predictors of performance (Röhlcke, Bäcklund, Sörman, & Jonsson, 2018). Earlier research differs in both methodology and approach, thus making it hard to compare (e.g., comparing games played in succession to a study that focus on cognitive tasks and questionnaires). It is therefore essential to further ask why and to what extent time on task is related to psychological studies of esports (Hulaj et al., 2020).

2.3.6 Learning Management Systems For Esports

A Learning Management System (LMS) is a program and/or interface to support teaching and learning through tasks (e.g., training, collaboration, deliveries, curriculum, exam results, attendance). Adopting an LMS in education gives students, teachers, and other stakeholders access to the learning and teaching environment without being restricted by time or distance (Chaubey & Bhattacharya, 2015). A learning management system is defined as:

> *"a software application that automates the administration, tracking, and reporting of training events"* (R. K. Ellis, 2009).

The educational environment benefits by the inclusion of an LMS by promoting stakeholder functionalities such as assessment, management, collaboration, and more (Chaubey & Bhattacharya, 2015). A digital subject such as esports can be well structured and managed if implemented in a LMS. To make sense of data that is produced within the esports discipline, implementation of a LMS can be seen as fundamental for esports analytics to be successful. Concerning research on learning and esports, most LMSs are not designed or capable to collect deep analytics and provide access to that data. Bodily, Ikahihifo, Mackley, and Graham (2018) identify three challenges with LMS for studying learning: 1) how to track learning that occurs outside of a LMS 2) limits that prevent real-time analysis 3) incapability of collecting click-level analytics. To solve these issues, developers of a LMS should incorporate software extensions or Application Programming Interfaces (APIs) to close these gaps. A Learning Management System should be the overarching technological structure, but should also provide access to other tools to further support learning and teaching.

Previous research has shown that a seamless integration of tools for a functional LMS is wanted (Bodily, Ikahihifo, et al., 2018). Students reviewed an analytical dashboard tool and criticized the potential of the tool because they wanted their online work summarized in one place (i.e., a LMS can be one "place", but offer several tools for use on the same platform). Concluding statements shows that a structured and centralized application is wanted to further motivate higher frequency of use and better sense-making of data (Bodily, Ikahihifo, et al., 2018). Learn2Esport is an example of a platform that tries to solve this issue of data sense-making, as described more in Chapter 4.

2.4 Learning Analytics

Learning Analytics (LA) is a multidisciplinary approach to studying learning and serves as a tool to support students in a learning environment with feedback or instructional content. By applying a multitude of methods (e.g., predictive analysis, algorithms, visualizations, relationship mining), one learning analytics purpose is to tailor the educational setting to an individuals needs and abilities (Avella, Kebritchi, Nunn, & Kanai, 2016). The First International Conference on Learning Analytics and Knowledge (LAK '11) defined learning analytics as:

> *"the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs"* (Long, 2011).

Freire et al. (2016) display a basic LA structure with integration in a LMS. Access to user- and interaction logs provides opportunities for in-depth analysis to identify patterns that correlate with some aspect that is essential to the field of the application (e.g., low material use and long periods between sessions may result in academic failure) Freire et al. (2016). With the continuous effort of digital advancements[7], it is normal for institutions to employ a variety of information systems (e.g., student diary, learning management systems, digital repositories) to capture data and the environment in which it occurs. This data can then be used to improve both student learning and school operations (Gaftandzhieva, Docheva, & Doneva, 2020; Varanasi, Fischetti, & Smith, 2018). One way for learning analytics to be beneficial for learners is through "learner awareness tools" and there have been multiple successful implementations of such tools in educational settings (Brun, Bonnin, Castagnos, Roussanaly, & Boyer, 2019; Macarini et al., 2020; Macon, Macon, & Phillips, 2016; Sancho, Cañabate, & Sabate, 2015; Varanasi et al., 2018). Tools of this kind can provide students with information about their learning activities, and presented as feedback through implementation in a LMS (Bodily, Kay, et al., 2018). Two relevant fields of learner awareness tools are Learning Analytics Dashboards (see Section **2.5.1**) and Open Learner Models (see Section **2.5.2**).

Previous research on LA revealed a gap in the focus of what data to collect and analyze, and the lack of effort to answer this question (Kivimäki, Pesonen, Romanoff, Remes, & Ihantola, 2019; Winne, 2017). Some research reports that learning analytics tools can be affected by which data is available instead of focusing on learning science, thus in worst case resulting in a non-impactful implementation of learning analytics (Bodily & Verbert, 2017a; Kivimäki et al., 2019). The most common types of data are 1) resource use (found in 75% of the articles), 2) assessment data (found in 37% of the articles), and 3) social interaction data (found in 34% of the articles). Less common (including all categories, not only data sources) categories, yet relevant for this research, are 1) text feedback (found in 18% of the articles), 2) manual reporting of data (found in 13% of the articles), and 3) usability test (found in 11% of the articles) (Bodily & Verbert, 2017a). LA normally uses automatically captured log data and combines this with other self-reporting data collection methods (e.g., interviews, questionnaires, surveys) to make sense of the data (Kivimäki et al., 2019).

[7] Following 2020 and the outbreak of COVID-19, distance education, gaming, and internet streaming has been essential. By March 2020, more than 1.38 billion students had to stay home, thus adaptation of online learning and learning management systems has become unavoidable (Brem, Viardot, & Nylund, 2020).

The aforementioned data collection methods that rely on self-reporting suffer from bias, distortion, and false memories due to the nature of human behavior, and may therefore not accurately represent a learners profile (Winne, 2017). Despite this, they do represent a learners beliefs (which alone can provide valuable insight), and is the closest insight one can get into motivation and meta cognition. These are important factors when considering self-regulated learning and the effect it may have on visualization of past performance and feedback. Although the aforementioned capture methods and analysis techniques can make LA mostly focused on automatic tool development, it is not as straightforward. Rather, human interpretation of data is essential to make sense of it, detect patterns and build models and visualization dashboards in a practical way (Bienkowski, Feng, & Means, 2012). Not only is human interpretation invaluable, but we must ask *why* we measure the data that we choose. Booth (2012) considers the importance of principles and propositions to successfully measure and analyze aspects that makes learning matter, instead of measuring whatever is available just because we can.

2.5 Interfacing With Users

2.5.1 Learning Analytics Dashboards

Learning Analytics Dashboards (LADs) are defined as:

> *"a single display that aggregates multiple visualizations of different indicators about learners, learning processes, and/or learning contexts"* (Schwendimann et al., 2016).

Dashboards for learning analytics are essential as they provide visualization of personal information and collect traces that users are leaving behind. These traces come from online activity, data logs and by self-reporting (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). Learning analytics dashboards are meant to support the improvement of self-knowledge and self-reflection through visualizations that help make sense of the data. If students are to act on their behavior and enact SRL principles (e.g., goal setting, monitoring, reflection), they must become aware of their knowledge and behavior (Vilalta, Giraud-Carrier, & Brazdil, 2009). By use of LAD, students may gain awareness and understand which behavior are most effective to develop their learning methods (Colthorpe, Zimbardi, Ainscough, & Anderson, 2015; Vilalta et al., 2009).

A learning analytics dashboard shows visualizations and feedback on data that were produced by subjects, thus providing beneficial foundations for personal growth and may prompt reflective actions to improve results and performance. Verbert et al. (2013) distinguish four stages concerned with learning analytics and learning analytics dashboards:

- **Awareness** - Concerned with data and visualization
- **Reflection** - Focus on users' asking questions and accessing relevancy
- **Sensemaking** - Concerned with answering identified questions
- **Impact** - Induce meaning or change behavior if concluded as useful

Learning analytics focus on learning and reinforcement of the learning environment (Avella et al., 2016). Assessment through information visualizations can help subjects gain insight into their data and trends, thus supporting reflection for self-regulation and reinforcement learning (Verbert et al., 2013). In a review of learning dashboard applications, Verbert et al. (2013) identify two variables to consider when developing a dashboard tool: target users (e.g., teachers, students) and tracked data (e.g. time spent, social interaction, document and tool use, artifacts produced, exercise results). As self-assessment and motivation are important variables to consider in a LMS (i.e., online learning in general), it is essential to let students know how they are doing. The teachers' role in this environment is to adjust the course, inform on performance, and identify students that are struggling (Freire et al., 2016). There are benefits for an LMS to apply LADs. Through studies and interventions of the captured data (both automatic and self-reported) it will be possible to adapt and customize the learner content to match a student's needs , but also enhance the role of the teachers and other stakeholders (Volk, Kellner, & Wohlhart, 2015).

2.5.2 Open Learner Models

In an intelligent learning environment (often as a learning management system) data on student's activites and performances are saved. The *learner model* is thus a recreation of a student's levels of knowledge and beliefs, and the functionality of a learner model is to adapt and produce individualized learning. Bull (2004) describe a learner model as:

> *"a model of the knowledge, difficulties, and misconceptions of the individual"*

Learner models are not structured for human interpretation, nor are they (generally) available for the students. Gaining access to the learner model may provide educational advantages and awareness of ones knowledge, thus *opening the learner model* can be an effective method for academic development Bull (2004).

Open Learner Models (OLMs) are defined as:

> *"making a machine's representation of the learner available as an important means of support for learning"* (Bull & Kay, 2010).

Bodily, Kay, et al. (2018) acknowledge that OLMs are created for multiple purposes, such as improving the learner model, supporting meta cognitive processes, facilitating navigation and decision-making, and addressing issues on personal data (Bull & Kay, 2007, 2016). Opening the learner model is not a simple task. The models are usually not designed to be interpreted by humans, and if they are, it requires some experties or domain knowledge to make sense of the data. A key challenge for OLMs is how to design interfaces to make sense of this data in a human-understandable format (Bull & Kay, 2010).

An open learner model in its simplest form, is a method to view the captured model of a learner, however, with no further functionality. This elementary approach is called inspectable learner models (Bull & Kay, 2010). This allows stakeholders to view the information available, despite not receiving any feedback or recommendations on how to process and proceed with the information. Inspectable models may raise awareness and cause reflective actions (Bull & Kay, 2010).

When opening the learner model, it is important to be aware of the outlined use: opening the learner model for a learner may have benefits as an inspectable model, but feedback, aggregated data and recommendations are also beneficial to promote self-regulated learning (Zumbrunn et al., 2011). In the instance of teachers, the inspectable learner model may be more powerful as teachers may effectively identify deviations, thus resulting in interventions. Concerning other stakeholders, such as researchers and regulators, the use of inspectable models may be more powerful as the data is as raw and unaffected as it can be, thus providing dynamic capabilities to aggregate data for assessment, and examine further trends that otherwise could be ignored.

There are also other variations of an open learner model (e.g., editable learner models, persuasive models, evidence based models, mixed control models) (Bull & Kay, 2010). The situation in which the learner model is to be opened, and to what purpose, should affect the approach (i.e., model choice), and how to perform visualizations or feedback in general.

In this research, there is a unique opportunity to assess game performance in a learning environment where there is a lot of data to handle and triangulate. Outlining the use of the learner model is important, and should be carefully considered when opening the model to effectively facilitate the motivation of opening the model in the first place. Previous research also suggest that learners apply higher trust in a system that offer a mixed control approach to the learner model (Ahmad & Bull, 2009; Bull & Kay, 2010).

The methods of adopting OLMs and building dashboards for intervention must be considered based on the context of use and the involved stakeholders. Bull (2004) arranged a survey for university students about their expectations regarding use of open learner models in the learning environment. Some of the results include: 1) 70% consider it their right to view their learner model, 2) 80% would use the model to reflect on learning, 3) 50% want an editable type of learner model 4) 73% find a mixed graphics and text learner model most appealing, and 5) 55% want to compare their learner models to their peers (Bull, 2004).

2.5.3 Opening The Learner Model

A survey conducted in 2004 had a total of 44 students that enrolled either in a course called "Interactive Learning Environments" or "Educational Technology" (Bull, 2004). This survey considered opening the learner model to the students, and captured their beliefs regarding such a system.

Bull (2004) accurately restricted her results in the sense that the survey participants may have more than average interest in the topic considering their choice of course enrollment. One must always consider that choices may be altered and opinions are changed when faced with a real example (and not a survey), and that the number of survey participants are too limited to generalize the results. However, the indicators found in this survey can be helpful guidelines when designing dashboard applications. By compiling the results and extracting the most relevant findings reveal some interesting conclusions. The majority of the students presume that opening the learner model: 1) is their right, 2) will help them plan and reflect on learning, 3) can be used as a navigation aid, 4) will add to the learner modeling process.

Concerning the different types of open learner models, the results were mixed. Half of the students had a preference for editable learner models, but these results are heavily influenced by uncertainty. This further strengthens the notion that a prototype iteration or application should be composed for the research phase. By omitting either a visual representation or detailed structural plans for a working application may confuse the subjects as they are left to their own interpretation of an open learner model. When asked about preferred content of the learner model, most students want: 1) access about their knowledge, difficulties and misconceptions, 2) either a graphical interface or mixed graphic and textual interface, 3) to compare their content to expert knowledge, requirements for course success, comparison to peers, anonymously presented to their instructor, and contribution to a class average (Bull, 2004).

2.5.4 LADs vs OLMs

The definition of LADs and OLMs does seem to overlap, and it might not be easy to differentiate the two. However, the two practices originate from different disciplines and have been developed mostly unconnected to one another (Bodily, Kay, et al., 2018). Because of this split, Bodily, Kay, et al. (2018) argue that a typical OLM tool is different from a typical LAD (and other student-facing dashboards) by focusing more on work within learner modeling or user modeling, whereas dashboards are generally established with a data-driven incentive with a scope bigger than just the learner model (e.g., goals, stakeholders). The authors also recommend that academics who focus on student-facing learning analytics should base their search on literature from both LADs and OLMs research, and recommends Learning Analytics & Knowledge (LAK) researchers to use the term *learner model* because of its relevancy and strong overlap. This can add to the fast growing learning analytics field (Bodily, Kay, et al., 2018).

2.6 Self-Regulated Learning

The development of student-facing dashboards should have pedagogical concepts and theories in mind. Jivet, Scheffel, Drachsler, and Specht (2017) state that *self-regulated learning* is the most commonly used concept for dashboard technology. Self-regulated learning (SRL) is defined as:

> *"a process that assists students in managing their thoughts, behaviors, and emotions in order to successfully navigate their learning experiences"* (Zumbrunn et al., 2011).

Despite what data is automatically captured as log data in LA tools, it can be very difficult to capture data about self-regulation in learners (e.g., motivation, meta-cognition). Generally, it requires a lot of effort from the students to engage in self-report programs related to their past performance, and it is not unusual to see a decreasing trend in participation (Kivimäki et al., 2019). The authors also identify the potential of structured learning diaries as they capture real-time interpretation of the learning processes (Kivimäki et al., 2019). The approach of this research will engage with unstructured learning diaries and investigate its correlation to SRL. Zumbrunn et al. (2011) visualize a cyclical model for SRL, See Figure 2. The model of Zumbrunn et al. (2011) is built up of three phases: 1) forethought and planning, 2) performance monitoring, and 3) reflections on performance (Pintrich & Zusho, 2002; Zimmerman, 2000). This model exposes the structure of SRL and serves it as an important foundation for this research to investigate esports in school and how to combine learning analytics and visualizations to foster self-regulated learners.

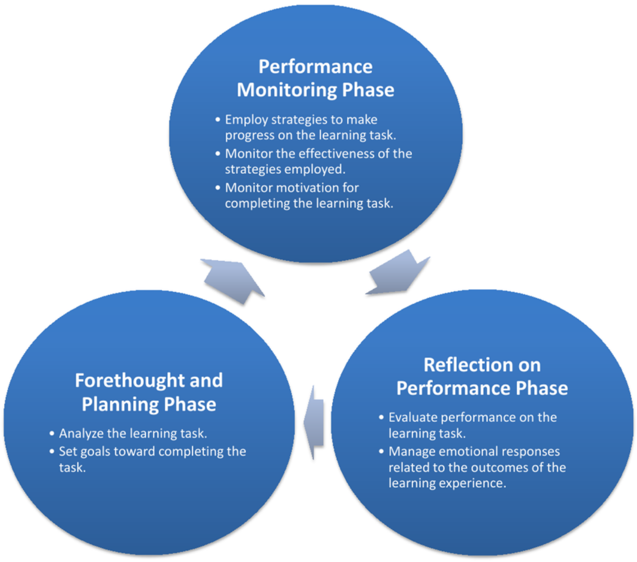

Figure 2: Phases of Self-Regulated Learning by Zumbrunn et al. (2011).

Table 3: SRL Processes, summarized from Zumbrunn et al. (2011)

SRL Process	**Process Definition**
Goal Setting	A goal may be determined as winning a game, completing a task, passing an exam, or gain understanding of a topic.
Planning	Complementary to goal setting, it is important to plan a schedule for how to reach the goals. Planning can be split into stages that include time allocation, specific task mastery and strategy choice.
Self-Motivation	This process would suggest that a student has control over their learning. By identifying learning goals, planning the learning process, and combining these with acts of self-motivation will make the learner more autonomous.
Attention Control	Ensuring the control of attention can be challenging, but is essential for SRL. Avoiding distracting thoughts and environments are important.
Flexible Use of Strategies	A student may have flexible strategies when working towards a goal. Young students are not always familiar with learning strategies, and it takes time to learn. Teachers and coaches play a vital role in this process.
Self-Monitoring	When a student wants to improve and acquire the other skills of SRL processes, it is essential to monitor their progress. Features of self-monitoring include, but are not limited to: time on task, attempts on task, and strategy use. Insight into these features help students visualize their progress.
Help-Seeking	When necessary, a self-regulated learner will seek help from others. This can be promoted by providing students with progress feedback that are easily understandable and effective.
Self-Evaluation	Students should evaluate their own learning independently (i.e., not only by teacher-issued assessments). This facilitates students to make adjustments to their learning processes

Zumbrunn et al. (2011) emphasize the importance of causal attributions (i.e., the student identifies causes for success or failure at a task) and its correlation to performance reflection. Self-regulation is often seen as a mindset or a lifestyle that must be taught for it to be effective; students need guidance, or must be taught self-regulated processes. Table 3 presents a summary and definition of key phases as described by Zumbrunn et al. (2011).

Depending on the task to be solved, self-regulated learners will modify their learning strategies and approaches, thus producing a learning environment that can be described as a cyclical and adaptive process to implement knowledge (Colthorpe et al., 2015; Zimmerman, 2000).

Broadbent, Panadero, Lodge, and de Barba (2020) commented on the opportunity that LA can bring to SRL through use of digital traces to better understand learning and the learning process. Instead of trying to explain learner actions based on outcomes, digital traces may provide researchers the opportunity to tie events and specific actions to SRL (Broadbent et al., 2020).

2.7 Games Analyzed

This section provides an in depth explanation of the video games that are played by participants in this research.

2.7.1 Counter Strike: Global Offensive

Counter Strike: Global Offensive (CS:GO) is the latest addition in a video game series called Counter-Strike. The game is a First Person Shooter (FPS) and is created by Valve (Valve, n.d.). The game launched late 2012, and has been one of the most popular esports games since its launch, and is still the most watched FPS genre-game (and second most watched overall) with 215 million hours watched by live audiences on broadcasting sites (e.g., Twitch, YouTube and Mixer) in 2019 (Newzoo, 2020). CS:GO's popularity does not seem to decline, but rather the opposite, as the average player counts have been steadily increasing throughout the years. The all-time-high record of people playing the game simultaneously was set in April 2020 with a staggering amount of 1.3 million players (Steamcharts, n.d.).

The game is played by two teams that play matches against each other: A Counter Terrorist (CT) team, and a Terrorist (T) team. There are with five players on each team, totalling ten players per match. Matches are played as best of thirty rounds, where each round is limited to one minute and fifty-five seconds. A round may be over before this time if every player on one team is eliminated. The terrorist team try to plant a bomb and make it detonate, whereas the Counter Terrorist team try to defend key positions named Bombsite A and Bombsite B. The bomb must be planted within a certain area on each bombsite, hence the Counter Terrorists always know where the Terrorists can plant the bomb, as do the Terrorists. The bombsites are marked on each player's radar and are static for each map in the game. If the bomb is planted, the Terrorists must defend the bombsite and assure the bombs detonation, and the Counter Terrorists must try to defuse the bomb (in competitive play, the bombs detonation time is thirty-five seconds, but has been prone to change). In rare occasions, the bomb may be planted at the last second of the round, boosting the total round time up to a complete maximum of two minutes and thirty seconds.

In order to win a round, the Ts either have to eliminate (i.e. reducing an opponent's health from 100 to 0) all the CTs, or successfully plant and detonate the bomb. On the other side, the CTs may win by eliminating all Ts, letting the rounds timer run out while minimum one CT player is alive, or by defusing the bomb. Because of these rules, there is a defensive incentive for the CTs to defend sites while the Ts must take the offensive initiative. There are many strategies to deploy from either side, utilizing key positions, communication, utility (e.g. smoke grenades, decoy grenades, flash grenades), audio cues and much more. After fifteen rounds, the teams swap sides and play as the opposite faction. In the case of a draw, as in each team winning fifteen rounds each, they go into overtime considering a best of six set. The overtime goes on until a victor emerges, and overtime on overtime is not unheard of.

At the start of each half, every player starts with $800, a default pistol, and a default knife. You earn money based on eliminations and round endings that reward the victors with a higher prize. Every player retains their money between rounds, and is only lost when equipment is bought, through penalty (eliminating yourself or your teammates), and at match half when teams switch sides. The economy has a lower limit of $0 and higher limit of $16000. When a player is eliminated, their main firearm, sidearm, and last equipped grenade is dropped (if you have not equipped a grenade, then the most expensive is dropped). Any player in the round can pick up the dropped weapon, and if they are not picked up they will disappear when the round ends. At the start of each round, there is a twenty second buy time. All items are bought within the spawn area, and one player can buy firearms to drop for their teammates. The first round of each half is called a pistol round. There is not enough economy to purchase other firearms than pistols, hence the name. Economical[8] management is an example of a tactic to employ. Good money management is essential and can often be a deciding factor for a match outcome.

The user interface of CS:GO provides many functionalities for players, and a lot of details to which players need to pay attention. Figure 3 highlights each section that is of importance. Player health and player armor informs the player of their health and armor amount (may differ from 0 to 100, but if health reaches 0 the player is eliminated). The image represents a player of the Terrorist faculty, and as they carry the bomb you see an image to inform the player of that. The players' economy shows how much money they have to spend on weapons and equipment. Map radar shows the blueprint of the map and identifies where Bombsite A and Bombsite B are located. A player can always see their teammates, and if the enemies make a lot of noise, the radar may pick up their location. The round timer, the amount of CTs and Ts alive, and match score is highlighted in the middle top of the user interface.

Killfeed is where information regarding eliminations take place. Player equipment showcases images of their equipment and which hotkey to access them. The bottom right shows the ammunition left in the current weapon, and total ammunition left for that round. The middle of the screen has a crosshair.

[8]The mentioned tactics and definitions are not final. These are results of continuous development and the current meta. Meta can be interpreted as an acronym for "most effective tactics available. A meta can change, and is traditionally a "community consensus" and not an official guide (Grammarly, n.d.).

Figure 3: CS:GO User Interface

2.7.2 League of Legends

League of Legends (LoL) is a Multiplayer Online Battle Arena (MOBA) game developed by Riot Games (Riot, n.d.-b). The game was released in 2009, and has since its launch been one of the most popular video games in the world. In 2019, LoL was still the most watched esports game (Newzoo, 2020). A regular match of LoL finds place in a map called Summoner's Rift where each teams side of the map mirrors the opponents side (See Figure 4). If you compare LoL to CS:GO, it is objectively a steeper learning curve to understand the mechanics and gameplay. The goal of CS:GO (or any shooter games for that matter) might be easier to grasp, and the gameplay is relatively easy to digest for the average spectator.

Due to the difference between an FPS game and a MOBA game, the official basic guidelines to LoL are summarized in Table 4. There is also a high diversity in roles to play in LoL. These are summarized in Table 5 with role names and role definitions (Riot, n.d.-a). Figure 4 is marked with some details of Summoner's Rift. **Your Base** marks the area that is considered your base. Here you find the Nexus, which is the final objective that must be destroyed to declare a winner. The first team to lose its Nexus is the loser. In the base you may also find the Fountain, and this is where you can buy items and/or restore essential character points. If you want to get to the enemy Nexus, you must clear a path. To successfully do this, your team must destroy defense structures, namely turrets and inhibitors. There are three distinctive paths towards the enemy base, namely **Top**, **Mid**, and **Bot**. Each of these paths are made up of three turrets and one inhibitor. Each team has five players, playing as a unique avatar that is referred to as a champion.

Figure 4: Summoner's Rift Map Overview

Each role conforms to a strategic play style, and is thus placed in a relevant position on the map. These are, as already mentioned, the three paths: 1) top lane, 2) middle lane, 3) bottom lane. A fourth role is a constantly moving player, and the fifth and last role plays a support role (Riot, n.d.-a).

Table 4: Basic overview of League of Legends Gameplay and Mechanics from Riot (n.d.-a).

Guide	Definition
What is League of Legends?	A team-based strategic video game with two teams that consists of five players each that fight to destroy each others base.
What do you play as?	Each player picks a champion to play for the duration of the match. One champion has special abilities and perks, and there are over 140 champions to choose from.
What is the goal?	Destroy the center of the enemy base - called the Nexus. The Nexus (and its surrounding area) is where you access the shop to buy items, replenish health, and is a spawnpoint for minions.
How to play towards the goal?	Clear a lane (i.e., a path) that blocks the road to the Nexus. A lane consists of turrets and inhibitors. To destroy these, you must deploy strategies that include teamwork and communication - the enemy team are trying to do the same to your lanes.
Other areas of the map	The jungle is separating the lanes, and this is where monsters and plants reside. Killing monsters and other units awards the player and their team with buffs to develop their characters.

Table 5: League of Legends Role Names and Definitions from (Mobafire, n.d.; Riot, n.d.-a).

Role Name	Definition
Assassin	High mobility and high damaging capabilities. Low defenses, and can be both physical or magical.
Fighter	Mixed versatility with mobility, damaging and defensive capabilities. Mostly physically oriented, but can also be magic or a mix.
Mage	Typically ranged and powerful damaging capabilities. Low defenses. Deals magical damage, and can be used both from range and close quarters.
Marksman	Deals consistent and ranged damage. Fragile defenses and dependent on their support for protection, but has high mobility. Deals physical damage.
Support	Excels at utility based gameplay. Protects allies, thus characterized with high defenses but low damage output.
Tank	Has high defenses and low damage output. Utility based gameplay, but with offensive capabilities. Disrupt enemies and create opportunities for teammates.

2.8 Related Research

This section summarizes previous research that studied esports analytics or similar approaches to learning analytics to contextualize preceding discoveries.

2.8.1 Multiple Datasets

R. A. Ellis, Han, and Pardo (2017) conducted a study on a class that used learning tools for tracking of observational data, and combined the findings with self-reporting questionnaires about their experience to gain further insight. Students had to interact with digital curriculum (in form of text, videos and, assessments). By combining learning analytics and theories of student approaches to learning, the study concludes its efforts of quantified insight, that the learning experience can be improved when combining observational data with self-reported student data (R. A. Ellis et al., 2017).

There are conclusions from previous research concerned with LA and the need to complement LA with other perspectives to properly assess the learning experience (Suero Montero & Suhonen, 2014). By examining the importance of emotion data in a learning environment, Suero Montero and Suhonen (2014) considered two methods of capturing such data: 1) through a system approach, preferably with use of a peripheral device, or 2) self-reporting of emotions. The paper identifies some promising scenarios for utilizing non-structured emotion data in a learning analytics scenario, being: 1) personal reflection, 2) monitoring of student well-being, 3) support, and 4) improvement of learner situation and course design (Suero Montero & Suhonen, 2014).

Pardo, Ellis, and Calvo (2015) looked into the comparison of results from quantitative data to results of qualitative self-reported data, and how they relate to one another. Pardo and colleagues conclude that the combination of multiple data sources can give insight not previously examined, thus discovering opportunities for new changes to learning design and curricilum construction (Pardo et al., 2015).

In another study, models were examined to combine self-reported data with digital traces stemming from data logs, and the same conclusions were reached; combination of numerous sources can provide novelty Pardo, Han, and Ellis (2016). Concerning online learning platforms, there is a need to develop analytical techniques that combine multiple data sources, namely self-report measures with other sources, to expands the fields of study (Gasevic, Jovanovic, Pardo, & Dawson, 2017; Pardo et al., 2015, 2016; Suero Montero & Suhonen, 2014).

Unfortunately, it is not as simple to say that self-report data exclusively will provide accurate and useful insight. There are well-documented flaws as to how to interpret such data, and how inaccurate the produced data may be (Colthorpe et al., 2015). Perry and Winne (2006) argue that self-reported data are not valid representations of either study activities or academic performance, but should instead be complemented with trace (or log) data that record the actual activities that occur. They do say, however, that self-report surveys are the most common approach to the measurement of SRL because it is efficient and can provide data about learners' perception of their learning process (Perry & Winne, 2006). If done correctly, this can produce trace data that accurately reports on a student's study behavior, and the self-report data can provide further insight to patterns and analysis to construct intuition only present when two (or more) data types are joined. Combining data types is essential, as stated by Colthorpe et al. (2015), where the authors study the combination of learning analytics and critical reflections. Self-regulated learning is developing from a uni-modal data focus (e.g., questionnaires, computer logs) to a discipline that embraces multiple data points. Colthorpe et al. (2015) identify a gap where the approach to such analysis still is bound to a single discipline (i.e., multiple data sources, however within the same structure, either only self-reported or computer logs), and call for research within environments with diverse data sources, such as a learning management system (Colthorpe et al., 2015). Using self-reporting technologies and analyzing data on behavior may reveal concepts of self-regulated learning that can explain the findings that come from combinations of multiple data sources. Perry and Winne (2006) claim that

> *"all learners are self-regulating but not all forms of profiles of SRL are equally effective (Winne, 1995), and not all learners are effectively self-regulating"* (Perry & Winne, 2006)

They call for further research to examine and understand student development and engagement with self-regulated concepts and how to capture them in learning environments (Perry & Winne, 2006).

2.8.2 Learning Theories and Dashboards

In a systematic review of learning analytics dashboards and use of learning sciences, Jivet et al. (2017) found that self-regulated learning is by far the most represented learning theory when concerned with dashboard design. More recent publications with theories grounded in humanism have focused more on 21st century skills where they put the learner at the center of the learning process (Jivet et al., 2017). Jivet and her colleagues also reviewed dashboard goals, thus trying to figure out the reasoning behind the choice of educational concepts. Four competencies were identified: meta-cognitive, cognitive, behavioral and emotional (Jivet et al., 2017). A fifth competency was also included, namely self-regulation, a competency that combines and involves all four identified competencies.

The authors observed, and to summarize relevant findings: **(1)** The majority of visualizations try to influence meta-cognitive competences in order to support awareness and reflection. **(2)** There is focus on evaluation of one's performance, and learning how to learn. This is also related to 21st century skills (Voogt & Roblin, 2010). **(3)** At the behavioral level, SRL concepts are also most commonly used. **(4)** The emotional level is mostly built on theories of 21st century skills (Jivet et al., 2017). Further on, the paper analyzed how information is contextualized in the dashboard applications, and identified three types: **(1)** social (comparison with peers), **(2)** achievement (goal achievement), **(3)** progress (compared to earlier progression). As an ending note, it is clear that self-regulated learning theory is the most frequently used foundation for learning analytics dashboard design, however, only 27.37% of the proposed designs have grounding theories based on learning sciences. The majority of the designs are based on data availability instead (Jivet et al., 2017).

2.8.3 Learning Analytics Competences, Expectations and Concerns

In the emerging days of LA, a survey was completed in the educational sector to measure the reception of LA and tried to form a common understanding of the domain (Drachsler & Greller, 2012). To benefit from learning analytics the subjects associated key skills (e.g., critical reflection of performance, digital literacy, and evaluation) to be critical for functional learning analytics. These skills align well with 21 century skills and SRL-techniques (Voogt & Roblin, 2010; Zumbrunn et al., 2011). The majority of survey participants specify that students need guidance and support to develop these essential skills. For future work, Drachsler and Greller (2012) stated that LA should focus on reflection support, and provide stakeholders with the necessary insight on their own performance. Revealing hidden information about the student (i.e., opening the learner model), was rated as the second most important objective to develop functional learning analytics (Drachsler & Greller, 2012). Dietz-Uhler and Hurn (2013) highlight some issues regarding LA, such as the role of pedagogy and the emphasis that pedagogy should initiate learning analytics, and not the other way around (Greller & Drachsler, 2012). Other concerns include: 1) discriminating successful and unsuccessful students, 2) adding data-driven expectations (i.e., expectations outside the scope of the learning environment and the faculty), 3) focusing on student retention and not student learning, 4) how applicable proxies are for learning (Dietz-Uhler & Hurn, 2013). The authors make clear that LA is not limiting learning if used correctly, but the approach to measurement and prediction must be clear and rational.

Verbert et al. (2012) studied the collection of educational datasets and proposed a framework for the analysis of educational datasets grounded in a researcher focused approach. The framework is built on three components: **(1)** Dataset properties, **(2)** data properties, and **(3)** learning and knowledge analytics (LAK) objectives. The inclusion of LAK objectives serve a purpose: Verbert and colleagues aimed to provide guidance concerned with the relevancy of any given educational dataset, based on disciplines and expectations of LAK applications (Verbert et al., 2012). See Table 6 for an overview.

Table 6: LAK Objectives by Verbert et al. (2012)

Objective	Definition
Predicting learner performance and modeling learners	Reveal variables that define the learner (e.g., performance and knowledge)
Suggesting relevant learning resources	Recommend relevant resources or actions
Increasing reflection and awareness	Visualize indicators of activity and results to increase awareness
Enhancing social learning environments	Network- and/or social network analysis to visualize social interactions
Detecting undesirable learner behaviors	Discover students with problems or difficulties meeting the learning goals
Detecting affects of learners	Detecting the underlying mood or emotion (e.g., boredom and frustration)

Bodily and Verbert (2017b) looked into issues and trends in student-facing learning analytics. There are many different terminologies for student-facing learning analytics (e.g., recommender systems, feedback systems, intelligent tutoring systems, educational data mining, or learning analytics) and venues for dashboard and feedback systems in general. Because of this, the authors approached their review with broad inclusions to maintain a representative consensus of the state of the art, and included previous research on student-facing learning analytics reporting systems (Bodily & Verbert, 2017b). The authors reported that 17% of the systems included both visualization feedback and recommendation segments, and state that further technological development should try to close this gap by including both functionalities in their systems. An additional finding was concerned with the aim of increasing trust. A system should provide transparent recommendations for further actions (i.e., contribute the reasoning behind a recommendation), as opposed to solely proposing a specific action. The technology should say *The proposed course of action is Activity A, based on Data N*, instead of *The proposed course of action is Activity A*. Bodily and Verbert (2017b) encourage future research to investigate the effects of additional information and multiple data sources into a student-focused dashboard application.

In a review of the state of the art of LA, it was shown that research in LA is data rich, but theory poor (Misiejuk & Wasson, 2017). Predictive analysis is a popular research area (e.g., curriculum issues), while predictive *models* are rarer as generalization is challenging. Generalization through context is difficult as the purpose of LA is to tailor education to an individual learner's needs through feedback, instructional content, or models (Avella et al., 2016; Scheffel, Drachsler, Stoyanov, & Specht, 2014). Some of the gaps in research on learning analytics are research on feedback, learning-centric analytics, and implementation of learning analytics (Misiejuk & Wasson, 2017). There are many data types and methods to consider in the field of learning analytics, and the approaches utilized are dependent on what data is available. For ***1) single datasets*** log data is the most frequent, ***2) two data sets*** show the most common pairing being performance data and log data, while ***3) three data sets*** most frequent combination is performance, log data and demographic data (Misiejuk & Wasson, 2017).

2.9 Research Context

Since esports takes place in the virtual world, every student action leaves a digital footprint that can be tracked, making esports different from traditional subjects. The technological development has expanded most school subjects[9] (in first world countries) to have a digital footprint, normally in a learning management system, but for esports the learning takes place in a digital environment for both learning and playing. This provides stakeholders with a unique opportunity to analyze data and perhaps understand learners in a way that has never been done. Identifying performance and development variables is not an easy task. The assessment of esports requires consideration of the following factors: teamwork, communication, solo-performance, game sense, overview, patience, self-control, timing, time spent, physical form, sleep, nutrition, exercise - the list goes on, and every factor may be an indicator of where something went wrong or right. This research does not aim to generalize the research findings, and is not concerned with the debate about "what does gaming have to do in education?". In this case, that matter is settled. The focus is how to create meaningful learning analytics with the available data in an effective way to: (1) compare performance data between user groups, (2) build a low fidelity prototype that utilizes the analyzed data, and (3) define enactment of SRL-principles in the data to study the differences of their usage.

The focus is how to create meaningful learning analytics with the available data in an effective way to: 1) open the learner model and present meaningful feedback to the students and teachers, 2) provide indicators as to the development of personalized esports curriculum, and 3) give other stakeholders and organizations better understanding of the learning process. As esports continue to grow as a discipline, it is important to provide young students with beneficial learning environments. Online platforms, or learning management systems, can provide deep insight, planning, and feedback to support improvement and learning. By further investigating esports in education and developing a profound understanding of data (e.g., data logs, activity data, game data, self-report data), it may be possible to discover new methods and innovations to foster self-regulated learners and advance curricular progress. Processes that define a self-regulated learner can be found in Section 2.6 in Table 3, and these are undoubtedly similar to the expected functionality any learning management system should have.

A vital aspect of learning analytics is the use of dashboards and visualizations for stakeholders to gain insight into the learner model. When developing dashboards it is essential to provide sensible assessment. Certain forms of feedback may prove valuable for some, while others can see it is a distraction. This supports the need for personalized learning analytics, favoring a tool with editable functionality. Limiting learning investigations to one-dimensional aspects (i.e., single data points or events) as representative of the learning process and learner performance is deficient. Considering learning management systems, it is important to include data necessary to advance the learning environment and develop tools that are crucial in such regards. In instances where the required or desired variables are either unavailable or non-existent, then *proxy* variables may be used. A proxy variable is a variable that is either used or calculated to represent an unobservable measure. Using proxy variables, as long as they are developed with close correlation to the variable of interest, can be powerful and helpful to reach the desired results or studies.

[9] For most subjects, the learning takes place outside of a learning management system. This produces a very limited amount of student data, making the captured data merely, at best, an oversimplified representation of a students learner profile.

At the same time, proxy variables can be dangerous, and in turn can set a precedent that produces invalid and unrepresentative results. Researchers must be cautious and persistent when applying proxy variable findings as they may be limited or invalid. Greller and Drachsler (2012) presented a guidance framework for LA development and tool expansions, with the following reminder

> *To judge a learner's performance merely on quantitative data, is like looking at a single puzzle piece*

Drachsler and Greller (2012) in both their aforementioned publications from 2012, can seem critical in their thoughts on LA as a discipline (Drachsler & Greller, 2012; Greller & Drachsler, 2012). For an emerging discipline it is important to be critical in the early stages, and the strict view and recommended future research presented by the authors still stand strong today. Studying learning in technology enhanced environments can prove challenging, but the focus should be on learner progression and with learning theories in mind, instead of getting excited by data availability and the notion of *tracking everything that is possible*.

3 Methodology

This chapter presents the approach to studying data that originate in an esports course that is offered as part of the curriculum in some Swedish high schools. By applying Learning Analytics as a research methodology (Long, 2011), the objective is to study esports data in detail to interpret meaning and reveal hidden patterns, and study the effect that self-regulated learning concepts have on activity and performance. A learning analytics process cycle is presented to guide the research in a cyclical nature where methods and objectives are explained for each component of the cycle.

3.1 Learning Analytics Cycle

The learning analytics process cycle is a standardization created by the subcommittee for *Information Technology for Learning, Education and Training*. The cycle is visualized in Figure 5, and is also described as:

> *"Standardization in the field of information technologies for learning, education, and training to support individuals, groups, or organizations, and to enable interoperability and reusability of resources and tools."* (ISO/IEC-JTC, 2013)

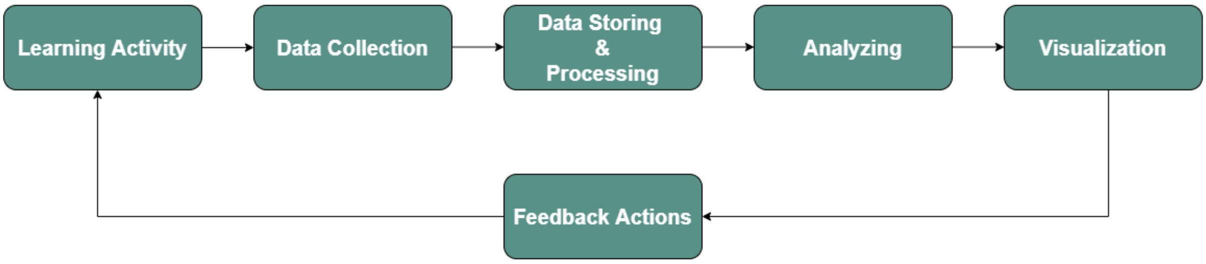

Figure 5: LA Process Model (ISO/IEC JTC1/SC36)

There are multiple stages in the LA cycle, and the components differ in length, size, and contribution to a research project. The model is meant to represent aspects of the LA process cycle to help guide research for information technology in educational settings. For this Master research, the following stages of the cycle have been used:

- **Learning Activity:** As part of *Desk Research 3.2* and *Exploratory Analysis 3.3*, the first stage of the model is concerned with getting to know the learning activity and the environment in which esports data is created. In this stage, exploration of the data and the esports management platform were the focus.

- **Data Collection:** The data was shared by the esports management system provider, Learn2Esport (Learn2Esport, n.d.). All shared data was anonymous with no identifiable properties. As part of the research team of eSports in Nordic Schools (eSportsNS), there is a collaborative relationship with Learn2Esport. Through discussion and testing, the relevant and needed data entries were queried and distributed for further processing and analysis.

- **Data Storing & Processing:** Following the data collection phase, the data was accessed and stored for further processing. This part of the cycle contributed to the overall quality of research, and assisted the preparation for a stable approach to the analysis phase by developing a plan for dataset measures, and how to: **(1)** clean the data, **(2)** sort the data, **(3)** identify, if present, faulty or invalid entries and how to spot them, **(4)** develop a plan for a dashboard prototype, **(5)** early progression and improvement in technical tools and practice, **(6)** develop and manage a dataset tool specifically for the data structure, and **(7)** iterate the unstructured data entries and develop coding categories for the self-report data analysis. This phase processed the students into user groups, developed proxy variables for activities, and prepared the data for analysis.

- **Analyzing:** After desk research, preliminary analysis, data collection, and data processing, the data was analyzed. Based on the data processing and user grouping, comparison between groups were completed, producing results as *Descriptive Statistics 3.4*, *Data Visualization 3.5 (trend analysis)*, and *Thematic Analysis 3.6*. Findings from different data sources were then compared and analyzed in combination through *Triangulation 3.7*.

- **Visualization:** For this research project, the visualization phase is part of the *Prototype Development 3.8*. After analyzing the data, a low-fidelity prototype was developed to provide an example for the presentation of student data in a student-facing dashboard.

3.2 Desk Research

The desk research (i.e., collection of secondary data) phase of this study included preliminary dataset analysis, an extensive literature review, background clarification, testing of data management tools, and getting to know the relevant games and game mechanics. In depth explanations of the query designs and databases used can be found in Chapter 2. An older dataset was shared for the purpose of preliminary analysis (or exploratory analysis). The complete dataset for 2020 was received early January 2021. On account of this, the desk research phase has been an important part in the development of capable groundwork. The preliminary analysis helped develop a grounded approach to limitations and rules for data analysis (i.e., algorithm development and testing) and journal coding (thematic analysis and content extraction).

3.3 Exploratory Analysis

Exploratory analysis is a broad method coined as *the act of looking at data to see what it seems to say* (Tukey et al., 1977). The goal of exploratory analysis is not to apply some model, but to look at the data from various points of views to observe, learn, and generate procedures for analysis and additional methods (Morgenthaler, 2009). This method was a major process in the preliminary analysis phase, creation of user groups, and opening of the learner model.

3.4 Descriptive Statistics

Descriptive statistics are used to describe the basic features of the data in the study, and is used to summarize information about data in a quantitative measure that can aid the process of definitions and analysis. Davidian and Louis (2012) said that statistics is the science of measuring, controlling, communicating and understand the data. A focal point of research in this thesis is the effect of self-regulated learning concepts and differences between distinct user groups. Using descriptive statistics can reveal information about students to describe their behavior and performance in relation to what the statistics reveal.

3.5 Data Visualizations

Data visualization is a common method for learning analytics (Misiejuk & Wasson, 2017). Visualizations uses lines, shapes, curves or other techniques to display relationships and patterns in the data (Manovich, 2010). In this instance the visualizations were used for *trend analysis* to identify patterns in the data. Different visualization methods can be used based on the data attributes, research vision and researcher preferences. Visualization of data can reveal insight and hidden patterns that otherwise is difficult to identify by working with raw data.

3.6 Thematic Analysis

As part of the preliminary dataset analysis and code book preparations, thematic analysis was used to find patterns and themes in the qualitative data (i.e., self-report journals). This method is widely used, and may describe datasets in detail. Thematic analysis helps researchers identify, organize and provide insight that can issue guidelines for further analysis or the identification of themes that are hidden or hard to detect (Braun & Clarke, 2012). Applying this method does not guarantee proper insight or answers to research inquiries. Rather, the findings that are revealed for the researcher must be relevant to answer specific research questions (i.e., the task for the researcher is now to identify the patterns that are relevant for his research inquiries) (Braun & Clarke, 2012). The utilization of thematic analysis was part of the desk research phase and helped produce a complete codebook with definitions and test coding before the final dataset was received.

3.6.1 Coding Data

The preliminary dataset was subject to thematic analysis, and is based on the following framework proposed by Braun and Clarke (2012):

- **Phase 1: Familiarizing Yourself With the Data** The first phase of this framework seem obvious, albeit necessary. Gaining access to a preliminary dataset provided an opportunity to know the structure of the data, and what to expect in terms of size, context, structure, language, and more. Taking notes, re-reading, or using tools are essential aspects of this phase - ensuring that the researcher thinks and reflects about what the data actually means. One of the aims of this phase is to start paying attention to details that are relevant for the research questions (Braun & Clarke, 2012).

- **Phase 2: Generating Initial Codes** The second phase pushes the analysis a step further, initializing the researcher to start coding. The codes are used as a summary of a specific data point. In this case, it is the journal entries written by the students.
- **Phase 3: Searching for Themes** The third phase is concerned with the transition of codes to themes. The depth of analysis concerned with the journals may prove to be a limitation: The code categories are synonymous with the themes to withdraw from the journals. The reason for this approach is simple: *If the student mentions Code A, then action X should take place.* As part of opening the learner model, data triangulation between the journal data to other student data can provide insight, maintain a flow of information about student progress, and also lay the groundwork to provide relevant feedback based on the journal entries. This method provides the most effective way to extract semantic meaning from the journals, hence the approach of coding category being identical to themes.
- **Phase 4: Reviewing Potential Themes** The fourth phase is an iterative process where themes are investigated through quality checking and redundancy. A descriptive statistical approach for the emerging themes were added to measure relevancy and to decide if the themes were appropriate or not. Through this phase, themes were removed, edited, or assessed as new themes to be added.
- **Phase 5: Defining and Naming Themes** The fifth phase is concerned with definition and naming of themes. As previously mentioned, the codes and themes overlap in this instance. The definition of each theme is structured in a way to explain how you can identify its occurrence in a student journal. This was also used in the coding process, acting as a supplement for the coders to use while coding.

3.7 Data Triangulation

The data in this research is a combination of log data, performance data, self-report data, and limited demographic data. The combination of two or more sources of data is called *triangulation*. By analyzing triangulated data, it is possible for the researcher to discover similarities between data sources to strengthen or validate research inquiries. Benefits of triangulation include: boosting data confidence, innovation, new understanding, finding unique patterns, and creation of a broader overview of the researched phenomenon (Thurmond, 2001).

3.8 Prototype Development

Based on data availability and the analysis results, a mid fidelity prototype for an open learner model dashboard technology was designed. Inclusion of a prototype served two purposes: (1) Examples of data visualization techniques for the development of a student-facing dashboard, and (2) visualize how data can be presented and used in combination for insightful feedback. The design for the prototype was based on the current design of the esports management platform created by Learn2Esport.

3.9 Technology

In this section there is an overview of the technologies that were used in this study. These include data processing (Google Spreadsheets) and prototype development (Proto.io).

3.9.1 Google Spreadsheets

Google Spreadsheets[10] is part of the Google Docs Editors Office Suite which is available as part of the Google Drive[11] service. The tool was initially released in 2006, but has since then received many updates and compatibility upgrades to support a broader use of the service. Google Spreadsheets is a platform that supports real-time query technologies (maintaining a structure close to that of SQL).

There are many built in functions[12] that can be used to manipulate data and execute calculations. Users of the tool may also access Google Apps Scripts[13], which is an environment to create custom functions and macros, store variables, and interact with Google Services. Google Apps Script is based on the JavaScript language.

The dataset (more about the dataset can be found in Chapter 4) was shared in .xslx format, which is supported by Google Spreadsheets. Prior experience led to the decision to use Google Spreadsheets as the platform and technology of choice for data analysis. There were iterations and testing in an environment with Python 3.9.2 and relevant packages (pandas, numpy, matplotib) for data analysis. Combining prior experience with success in preliminary analysis, immediate execution, revision history, and flexibility, the query-like tool from Google became the superior choice for this project. See Appendix A for a full overview of the queries and functions used in this research.

3.9.2 Proto.io

Proto.io[14] is a prototype development tool that supports the design of interactive prototypes. The tool offers a wide library of pre-made templates and icons, and also support for multiple design dimensions and operation systems. The decision to include proto.io as the prototype development technology was for its easy-to-use interface and the option to host (i.e., run) the prototype as a semi-functioning interface.

[10] `https://google.com/sheets/about/`
[11] `https://google.com/intl/en_in/drive/`
[12] `https://support.google.com/docs/table/25273?hl=en`
[13] `https://developers.google.com/apps-script`
[14] `https://proto.io/`

4 The Dataset

This chapter gives an overview of the dataset and its properties used in the research. First, the data provider is presented. Second, data processes are described, which include a dataset overview, how the data was prepared, user grouping, proxy development, and the process of coding qualitative data.

4.1 Ethical Concerns and Consent

The analyzed datasets were anonymized by Learn2Esport who shared the data for this thesis. No personal information was involved in the process. None of the data points contain sensitive information about the user that can lead to identification. This covers both real personal information (e.g., name, age, gender) and online persona (player profile names). The names of the participating schools are also made anonymous. As part of Esportsns, this project and the involved users have been employed as researchers within Learn2Esport, and non-disclosure agreements have been signed for confidential disclosure. The research is completed for the purpose of development and to gain knowledge about functionalities and trends in the esports programs for Learn2Esport and how to handle data from an esports environment.

4.2 Participants

The subjects in the dataset are high school students from Swedish schools that offer esports as a course through Learn2Esport. The majority of students are male, and they are between 16 and 19 years of age (Learn2Esport, n.d.).

4.3 Learn2Esport

Learn2Esport was founded in 2016 by sports and esports experts and enthusiasts to support the educational needs of students (Learn2Esport, n.d.). They create educational content to support 21st century skills - such as critical thinking, collaboration, entrepreneurship, and creativity - and they deliver this as a solution in an esports management system: Gameplan (Gameplan, n.d.). This is a one of a kind learning platform with solutions for stakeholders to manage esports programs, teams, camps and centers. The platform makes it possible for teachers to create and manage lessons and training programs, and it also gives teachers insight into player performance and development. Students gain access to lessons in form of theory to read, drills to complete, and exams to pass. Some of the other functionalities include player profile linking, journal writing, and game reviews (Gameplan, n.d.; Learn2Esport, n.d.).

4.4 Dataset Properties

The data is built up of multiple components that jointly represent the learner model and all the available data associated with each student. The data can be split into separate modules that gives an overview of the relationship between different data properties, as well as components within the distinct data sources. Through triangulation, the data can be seen as one entity, and can produce unique findings that otherwise are not available. To effectively open the learner model and present the data in meaningful ways, it is important to understand concepts of learning and causal attributes in the data. Figure 6 is a visualization of all available data properties within the dataset.

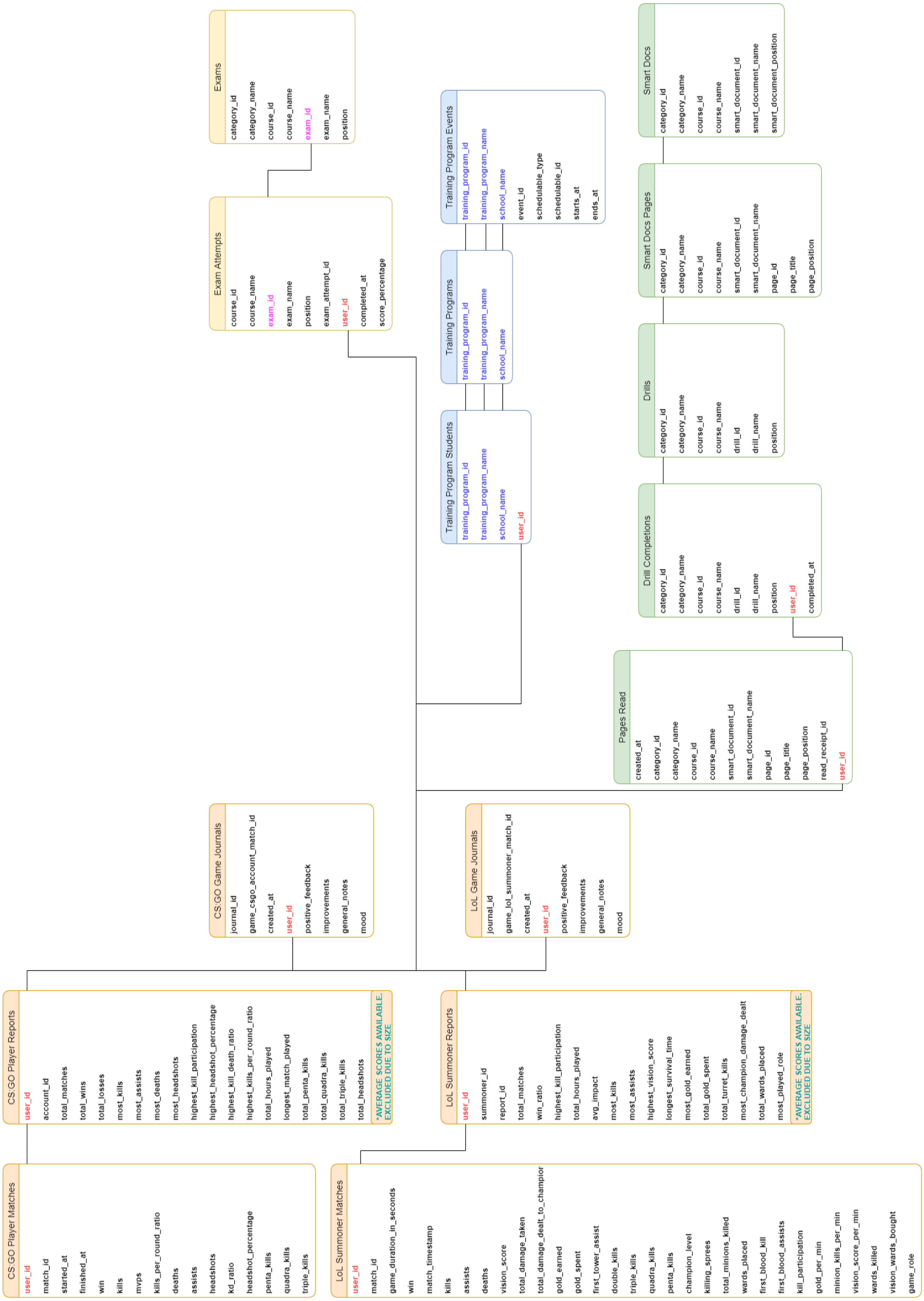

Figure 6: Dataset Overview

4.4.1 Student Info

Student activity on the platform are tracked and stored as activity data, or trace data. Activity data can be split into three components: **Theory Activity**, **Drill Activity**, and **Exam Activity**.

Theory activity gives an overview of what theory pages and lessons the students have read and interacted with. This includes a timestamp to see when the students accessed the data (but not for how long they stayed on the page). Data properties on theory activity includes what topics have been read (name of the pages), the overarching categories (what lessons they belong to), and what game they belong to (either CS:GO, LoL or General Courses). **Drill activity** is compiled with drill activity data for students. The data have timestamps to see when the students finished a drill. Data properties related to the drill activity entries include the lesson that the drill belongs to, what game it belongs to, and the name of the drill. **Exam Activity** has data on the exam attempts made by students. This includes what category the exam was about, when it was completed, the exam results (as a percentage), and if the student completed the exam or not (i.e., if the student quit the exam, then there are no timestamps or percentage scores).

4.4.2 Self-report Data

When a student has completed a match they have the option of self-reporting on how the game went. To make the self-reporting functionality available, the student must have connected their player profile to the LMS. Both processes of linking a player profile and writing a journal are optional actions. When self-reporting on performance the student can provide information related to multiple aspects of the game: *What can they improve on, what did they do well, and general notes*. To supplement the match feedback, the students can also rate their mood (on a scale from 0 to 4) that quantifies their perception of their mood throughout a match. Data properties include date of journal entry, user ID, a game match ID, unstructured data inputs, and the reported mood. There were a total of 82 unstructured journal entries to consider, written by 16 unique students.

4.4.3 Exam Performance

Every active student (i.e., a student that currently enrolls in an esports course) are part of training programs at their respective schools. Based on this training program, the students get a calendar overview of what topics to read, what drills to complete, and also when to have exam assessments. The exams are built upon previous experiences within the platform and are related to the esports game they are playing, and esports topics in general. An exam in this instance is not related to a final grade for the course (the course does not produce a final grade), but exams are instead used as a measure of knowledge that produces a percentage score for that said exam. The data produces insight as to when the student attempted an exam, the score percentage, what overarching topic and course the exam belongs to, and if the student passed or not.

4.4.4 Game Performance

One of the functions of the platform allows students to link their player profile to the platform. This ensures that all games played on the linked accounts are tracked and can be used for deeper analysis.

It is then possible to compare match performance while enrolling in the esports program to their general statistics (given that they have played the games beforehand). The aggregated data on the player profiles are data rich. This data includes identification of the user ID, some common statistics between the two games (e.g., total matches played, total hours played, total wins, total solo performance aspects and more), and game specific details. For a full overview of the data properties that can be found in this dataset check Figure 6. If the students have linked their accounts their matches will be tracked[15]. This dataset includes data for specific matches totalling 24206 entries from 63 unique students for the year 2020.

4.4.5 Demographic Data

The remaining data create a demographic overview over the training programs that students are admitted to, and what schools they go to. A total of four distinctive schools are included. There are no demographic data on the students (i.e., age, gender), but the students can be in the age range of 16 to 19, and the majority of the students are male (Learn2Esport, n.d.).

4.4.6 Data Flow

Figure 7 provides an overview of the data flow for the LMS processes, and its relation to the LA process cycle.

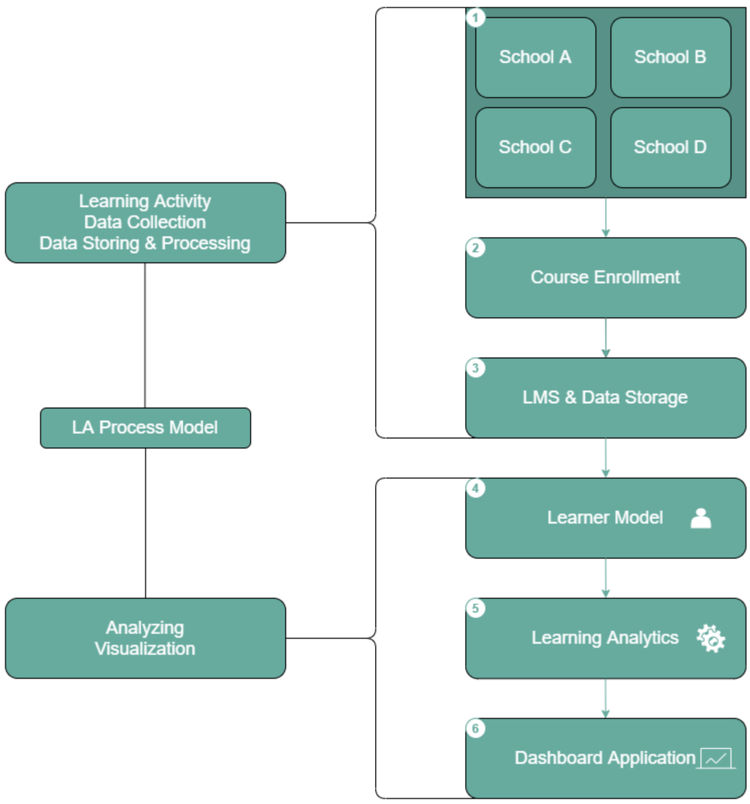

Figure 7: Data Flow Overview and Relation to the LA ISO Model

[15]Both games included in this research requires player profiles where users have to sign up with usernames, passwords, and emails. These profiles may be open to the public or access may be gained through other vendors. The esports management platform offers a way to import profile data, and that functionality is labelled *profile linking*.

(1) Students attend a school, and **(2)** enroll in the esports course. As part of the curriculum, the schools utilize a **(3)** learning management platform (Learn2Esport) for their esports content and student participation. The first three stages are related to the *Learning Activity, Data Collection, and Data Storing & Processing* stages of the LA process model (Figure 5). The activity data is based on student interaction with the LMS and match performance is provided through third party technology integrated to the platform. Data collection is automated within the learning environment, and is only limited by the choices that students make (i.e., if a student does not connect their player profile to the platform, no match data will be stored). Access to anonymous data was provided by the platform both for storage and processing.

Performance data and activity data from the LMS make up a **(4)** learner model for a student. This data can be used in **(5)** learning analytics processes. Analysis of student data can develop (or upgrade) a **(6)** student-facing dashboard application to provide insight, visualizations, recommendations, and other functionalities. The purpose for this is to help develop self-regulated learners, or promote self-regulated learning concepts as effective tools for learning. These steps are related to the phases of analysis and visualization of data in LA processes.

4.5 Data Processing

The preliminary analysis phase contributed to effectively identify entries that could be considered invalid or unusable for the research. This included empty and/or faulty entries, empty date entries, invalid journal entries (i.e., if the written input is empty, not concerned with the game, not serious). The process of identifying invalid journal entries was an iterative process (See Table 7). Inconsistencies in the dataset had to be considered and were quickly identified to exclude any faulty analysis or misinterpretation. The data was sorted by date for a natural time based approach. Enrollment in the esports courses may span for more than one semester. Consequently, some of the students have data that span for more than one semester. Based on the dataset and structure of the course, there was a natural limitation in terms of data inclusion for all of 2020. Older data (i.e., data that pre-dated 2020) was beneficial in an iterative manner to practice the coding and translation processes.

Certain functionalities on the esports management platform were not fully operational at the start of 2020. This created some limitations in the data in terms of trend views (e.g., there is no data for drills or exams pre-dating August 2020).

Table 7: Iterative Process of Preparing Journal Entries

Iteration Stage	Iteration Action
Iteration 1	Label invalid entries (empty, out of context, not serious)
Iteration 2	Translate from Swedish to English and remove invalid entries
Iteration 3	Clean the journal entries: Removal of redundancy and generalization
Iteration 4	Redo Iteration 3 after finalization of the Codebook
Iteration 5	Create new dataset for inter-coder reliability check

4.5.1 User Groups

The first stage of data processing considered the students and how to group them based on SRL-concepts and data availability for analysis (See Figure 8). The following user groups were identified:

- $N_{Total=149}$: The total number of unique students in the dataset.
- $N_{PL=63}$: The number of students who have linked their player profile to the platform. *PL* is short for *Profile Link*.
- $N_{NPL=86}$: The remaining students after Total minus PL. These are the students who did not link their profile. *NPL* is short for *Non-Profile Link*.
- $N_{JE=23}$: The number of students part of Profile Link, and who also have written one or more journal entries. *JE* is short for *Journal Entry*. This group is a subset of N_{PL}.
- $N_{NJE=40}$: The number of students part of Profile Link, but who have not written any journals. *NJE* is short for *Non-Journal Entry*. This group is a subset of N_{PL}.

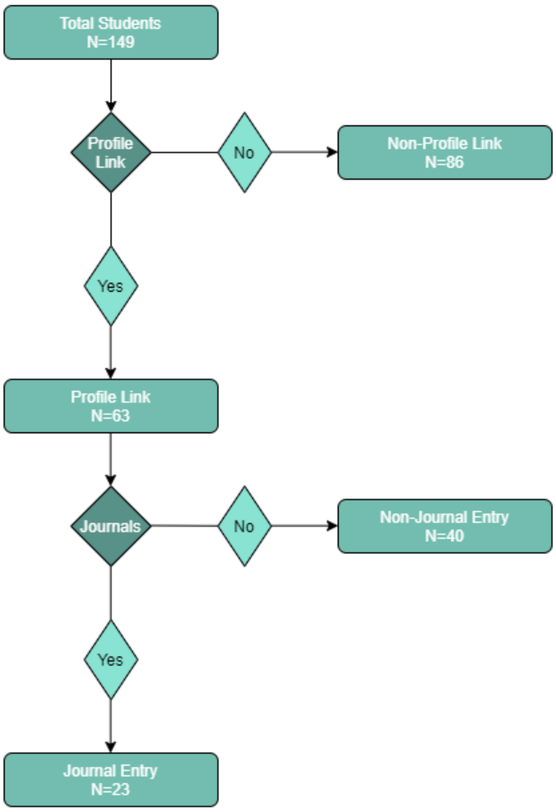

Figure 8: User Grouping Flow Chart

The five groups reflect different aspects of the data properties that are available, and the groups were identified to differentiate performance, activity, and other measures based on their enrollment preferences and activity. The dataset has a total of 149 unique students who enrolled in the esports course in the year 2020. It is voluntary for a student to link their player profile to the learning management platform. This creates a split in the data where there is more data on students who linked their profile, compared to a student who did not link their profile.

Out of the total 149 students, there are 63 students who linked their profiles. These 63 students were grouped into N_{PL} as **Profile Linking** students. The remaining students that did not link their profiles were then grouped into N_{NPL} as **Non-Profile Linking** students. Further analysis of the data for N_{PL} students revealed another layer for user grouping; performance reflection in terms of journal entries. Students that **Linked their profile** and wrote **journals on their performance** were defined as N_{JE}. The rest of the students that **linked their profiles**, but did not write any journals were grouped as N_{NJE}. Both of the groups N_{JE} and N_{NJE} are subsets of N_{PL} and do not count as new students.

As previously mentioned in Chapter 2, there are three phases of SRL visualized in a cyclical model (Zumbrunn et al., 2011). The three phases applied here consisted in 1) forethought and planning, 2) performance monitoring, and 3) reflections on performance. Group N_{PL} can be described as **performance monitoring** students. When the player profile is linked to the platform they gain access to aggregated profile statistics and past match performance. Group N_{JE} can be described as **students that reflect on performance by writing in the journal**. All students (N=23) that reflect on performance are also part of group N_{PL}.

Three comparisons were conducted: **(1)** comparison of students who linked their profile N_{PL} vs students who did not link their profile N_{NPL} **(2)** comparison of students who used the journal functionality and linked their profile N_{JE} vs students who did not use the journal functionality, but still linked their profile N_{NJE}, and **(3)** comparison of students who used the journal functionality and linked their profile N_{JE} vs students who did not link their profile N_{NPL}. Further comparisons are redundant as N_{JE} and N_{NJE} are subsets of N_{PL}.

4.5.2 School Enrollment

Considering ethical guidelines the names of the schools were removed. There were a total of four schools that the students attended. Some students were marked with *No School Enrollment* (i.e., the school entry is empty). Once the students concluded the course they were removed from the platform by the admins or coaches. Therefore, when the data was queried and extracted, some students had already been affected by this. To cover this issue, there was a fifth entry created for the school enrollment statistics labelled "No School Enrollment".

4.5.3 Theory Activity

Data on theory activity included duplicate values (i.e., if a student had visited any theory page more than once). Duplicates were not excluded because routines and study techniques will vary between students, and reading theory more than once is allowed. The total number of students was 149 in 2020, although three students had no theory entries. Therefore, the total unique entries of theory interaction for all students (N=146) was 11668. Despite having no theory interaction in the data (N=3), all three students had their profiles linked and generated game data on themselves.

4.5.4 Drill Activity

The drill activity entries included duplicate values (i.e., if a student has started a drill more than once). A drill may be completed as many times as the student wants, and drill entries with no date value were defined as a failed or aborted attempt (i.e., entries may be started, but are not always completed). These entries included the student id, but no date entry, and were measured as Non-Completed drill entries. Total drill entries equals 715 that were completed. There were 776 entries in the data, but 61 of the drills did not have timestamps, which means they were not completed. The total number of students in the drill data equals 72 unique students.

4.5.5 Exam Activity

Exam activity has information on the exams that students complete on the platform. There are four scenarios to consider:

- **No completion date an no score percentage:** The exam is started by the student, but not finished.
- **Completion date exists, but not score percentage:** The exam has been submitted, but has not yet been graded.
- **Completion date and score percentage exists:** The exam is submitted and graded.
- **No completion date, but there is a score percentage:** This is an invalid scenario and is caused by a technical error.

The number of graded exams is irrelevant for analysis as they rely on automated computer assessment or the efforts of a teacher. A total of 242 exams were started (this is the total count of exam attempts in the dataset, ignoring invalid attempts). There are four attempts considered invalid, as they have *no completion date, but there is a score percentage*. These entries were removed from the dataset. The remaining 238 exam attempts were analyzed. The exam results could range between the following values: 0%, 33%, 66.67%, and 100%. To pass an exam, a student must score 66.67% or higher, and 100% equals a perfect exam score.

4.6 Proxy Variables

Calculating a proxy variable can produce a representation or estimation of some variable if data is missing or non-existent. Proxy variables are not direct representations, rather, they can be used as indicators for further analysis and comparison. Following are description calculations for proxy variables for: time spent on theory, experience rate, monthly deviating win rates, and recommended session lengths.

4.6.1 Time Spent on Theory

To understand the development of the proxy variable, a **block** was initialized as a unit of measure. A block is defined as *a measure where the date and time stamps are considered to be from an identical activity session*.

Consider any click (i.e., page visit) on a theory page. An entry has a date and time of visit (format dd:mm:yyyy hh:mm:ss) when a student accessed the page. This is identical for any variable n in the data related to theory interaction. There is no saved data for when the page was closed or when the student moved to a new page. To create a measure of the time spent on page, the time difference between two sequential entries were. Consider the series $\{x, y, ..., n\}$ where each variable represents the timestamp for any given entry sorted sequentially. The time difference between x and y was calculated by

$$y - x = z \tag{1}$$

where x is the timestamp for any entry in the dataset, y is the timestamp for the subsequent entry to x, and z is then the time spent on page x. To maintain a rational variable calculation, the algorithm[16] had a maximum limit of five minutes for variable z to consider entries x and y to be within the same block. There are no upper limits in terms of the count of entries within any one single block. The lower limit is one. Any block m must have a minimum number of *one* entries to be considered a block. The only upper limitation that defines the end of a block is marked by a z value of five minutes or higher. The blocks were distinguished to get an overview of the student activity frequency in terms of time that has been spent working with theory and lessons. See Figure 9 for a visualization of the block structure. If a theory entry exceeds the time limit of five minutes it will define the end of the current block N^M and will move on to the next block N^{M+1}.

All entries that span five minutes and entries with no time spent value (i.e., the last entry in the data for any given student; the value cannot be computed if there are no subsequent values to any entry n) were excluded from the calculations. Consider student k. The following steps were taken to exclude entries considered invalid for a time spent on theory calculation for student k.

- **Step 1:** $\sum AllTheoryEntries$: Sum all theory entries the student had during 2020.
- **Step 2:** $\sum AllEntriesExceedingFiveMinutes$: Sum all theory entries that were calculated with a time spent value larger than five minutes
- **Step 3:** Subtract the answer from **Step 2** from the answer of **Step 1**

Distinguishing entries that exceeded the five minute limit created an opportunity to calculate total time spent values and average time spent values per student without being affected by: **(1)** the last page that was visited during a session. The time spent value for the last page can in some instances span for days or even weeks if the student did not interact with theory for a while. **(2)** generally any entries exceeding five minutes. If a student is away while the page is open or if the student does not pay enough attention to their task at hand. The exclusion of these entries was done to provide an accurate measure of time spent working with theory. The proxy variable is not meant to be a realistic value of time spent, but rather a representation because the desired values or calculations were not available.

[16] *Algorithm* and *query or queries* are used interchangeably when describing the methods for processing, analyzing, and working with the data.

One issue to consider was the query development in relation to date differences. The algorithm would flag a date difference despite a time difference of less than five minutes (e.g., if one entry occurred at 23:59:00 and the next entry was 00:00:01 the next day). This was controlled by querying all date and time entries, which revealed the latest entry to be at 23:43:06. If the theory interaction would measure over midnight with that start time, then the duration would exceed the five minute limit, and be flagged despite of the date change.

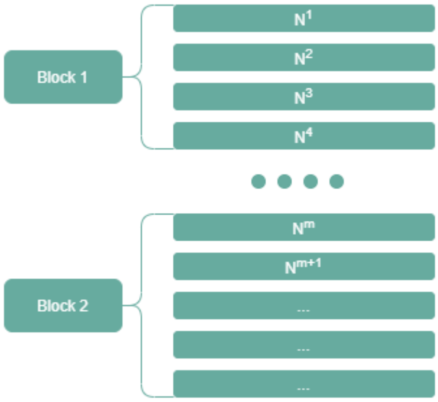

Figure 9: Time Spent on Theory Block Structure

The choice of five minutes as a limit is a result of the preliminary analysis where data on theory entries were analyzed from prior entries from the year of 2019. The preliminary analysis revealed an average time spent on pages of 21 seconds. Considering a vastly larger dataset for the 2020 data, and to account for variance in study strategies, the five minute limit per page was set. Other time limits were also tested (three minute, seven minute and ten minutes limit) but the results were not significantly different, thus it was decided that five minutes was to be the time variable for the proxy development.

4.6.2 Experience Rate

Analysis of experience and match performance was split into sections; one for CS:GO and one for LoL. The data from the students in these analyses were those of group **Profile Link** (N_{PL}=63) as performance analysis is related to game performance, and only the profile linking students created data on their gameplay. Further grouping within these students were made based on journal entries. In both of the sections there were comparison measures based on the students' performance between two distinct user groups: N_{JE} (i.e., students who used the journal writing functionality and linked their profiles) and N_{NJE} (i.e., students who did not use the journal functionality, but linked their profiles. The students could have played both games, therefore the same individuals can be among the data of both CS:GO and LoL.

There was data about the total number of hours played per profile, but this does not only include playing time, and is thus a bad indicator for time spent. Instead, an indicator for time spent will be the number of matches played. A user may be away from the computer while the game is running. Further analyzing this, results proved that a single user represented 91% of the total hours played of the group despite only accounting for 60% of games. It is reasonable to state that *time spent in-game* is a bad indicator for time spent *playing the game*. Thus, the data on total hours per profile was not used in the analysis.

An *indicator* for time spent was measured as the number of matches that has been played, while a proxy variable for an *experience rate* was calculated. The experience rate is a value (presented as a percentage) to calculate to what extent the number of matches played during enrollment account for the total number of matches played on that profile. Experience rate was calculated by

$$\frac{\sum EnrollmentGames}{TotalGames} \times 100 = Experience\,Rate \tag{2}$$

where *EnrollmentGames* is a count of all matches played during 2020 enrollment, and *TotalGames* is the total number of matches played. Having an experience rate of 100% would mean either: (1) the student is a new player, and has no prior experience (2) the profile is new and has no prior matches played before enrollment. It is impossible to determine if a student has prior experience by other measures (based on the available data), thus the experience rate is treated as a proxy variable that represents the experience level of a subject. A student with an experience rate of 50% would indicate that half of all matches were played during enrollment. The lower the experience score (can be between 0% and 100%) the more game experience the student had *before* course enrollment. By analyzing the data, comparing between groups, and interpreting results, it can be useful to look into how *experienced* the players are in terms of number of matches played, and compare differences between students that are *assumed* to be less experienced or more experienced. Consider the following example:

- **Student A** is part of N_{JE}, has an experience rate of 67%, and a win rate of 54.23%.
- **Student B** is part of N_{NJE}, has an experience rate of 32%, and a win rate of 51.17%.

Student A has the highest experience rate, and is thus *less* experienced between the two students. Despite a lower experience rate, the student shows a win rate that is higher. The experience rate can reveal differences that can explain choices made by students: *Student A has an experience rate of 67%. The student has therefore only played 33% of their matches prior to the course. As this student can be considered somewhat new to the game, they may be more motivated or in need to learn, thus their usage of platform functionalities is understandable. The student is part of the group* N_{PL} *and has therefore written journals about their game performance. This student shows a higher win rate compared to Student B who did not write game journals.*

4.6.3 Monthly Deviation

The monthly win rate for students were calculated. After producing an average win rate for the full year (win rate for enrollment), an algorithm flagged the months with the largest deviations from the average win rate (i.e., compared to the average win rate, the month with the highest win rate was flagged, and the month with the lowest win rate was flagged). The month with the largest positive deviation was labelled *positive month*, and the month with the largest negative deviation was labelled *negative month*. The purpose of finding the months with the largest deviations were to: (1) look into the number of matches played, (2) how many unique dates had match data, (3) how many games were played in each session. Findings from the months were then compared to effectively differentiate trends, frequencies, and session lengths to be able to recommend to a student *how to enact* their months with the highest win rate during 2020.

A third month was also included in this calculation; the most active month, as determined by the number of matches played. If the most active month was already labelled as the positive or negative month, then the second most active month was chosen, and so on. An average month was included to compare how the average trends and activities of a student looked like, and how different the deviating months were. A month with bad performance can describe something about the student that did not work, and a positively deviating month can focus on investigating habits that work better.

To ensure an adequate number of matches were played in the deviating months, a limit of *minimum* fifteen games was chosen. Example: *If a student has only played two matches in a month, and won them both, then the win rate would equal 100%, and that month would be labelled as a* positive *month.* Investigating trends and frequencies for only two matches would not reveal anything significant, and could certainly not produce recommendations or generalizations for future action. It is also difficult to determine an average match length based on multiple factors (e.g., rank, stalling, overtime, technical problems), and considering that the esports enrollment is in a school environment, a realistic lower limit of matches per month should be used. Fifteen matches can equal anything from 7.5 hours to 22.5 hours, which is a reasonable limit[17] to consider for a month to be representative of a student's gaming trends.

4.6.4 Session Lengths

Individual performance and session analysis included the top five most active students from each of the games. Students were filtered on how many matches were played during the whole year, and the top five students were chosen from each game (N=10). It was interesting to focus on students with the highest number of matches and top performance since it is expected that these students were the most engaged when playing, and they were maybe the most engaged in using the platform. There were multiple justifications to limit the analyses of session lengths to these ten students: (1) two games were included in the research, and this approach includes both, (2) many students didn't reach the lower limit of 15 matches per month, (3) the inclusion of esports as a course is new, thus functionalities on the platform, as well as procedures of course management are still in development and can be considered new. Thus, it was decided that focusing on students with a higher number of matches was more valuable

When the *deviating months* were identified, the matches for those months were sorted into groups. The match grouping was done to get an overview of the number of matches played *in each session*. A session has the same fundamentals as a *block* (Section 4.6.1), and is defined as *a measure where the matches are considered to be from an identical gaming session*. All matches have a timestamp when the match started. CS:GO (format: yyyy-mm-dd hh:mm:ss+00) and LoL (format: mm/dd/yyyy hh:mm:ss) had different time formats. Thus, a new dataset was created with a common date format (format: dd:mm:yyyy hh:mm:ss) to work with the data. Data on matches does not contain information about when the match ended.

[17] Can be compared to expected time to spend on homework for a course, or the total number of hours people should stay physically active per month. The most important functionality of this limitation is to negate the effect of luck, randomness, and other issues that may occur with smaller population samples.

In the same format as calculating time spent on theory (Section 4.6.1), the match length was computed to measure the length of a match. To consider matches to be from the same sessions, the time difference between match start had to be less than two hours. Example: *Consider the matches: Match A and Match B, where Match A was played before Match B. If the timestamp of Match B is less than two hours after the timestamp of Match A, then the two matches are from the same **Session**.*

The choice of two hours as a limit is a result of the potential match lengths for both games. The maximum length of a match (for both CS:GO and LoL) can reach up to one hour and thirty minutes, but such occurrences are rare. Keeping maximum match length in mind, and considering other events (e.g., short break, queue time), a limit of two hours was set to consider matches to be from the same session. Estimating the number of matches in a session was done to gain insight into student trends and what session lengths may be considered better in terms of performance (i.e., winning a match).

Feedback on *recommended session lengths* should be somewhat reasonable in terms of what is both *realistic*, but also considering aspects of *health, time management, and excessive video gaming*. Managing such aspects are out of the research scope, but reasons must still be given for the chosen limitations: Consider an average match time between 45 minutes and 60 minutes. *If a student is recommended to play sessions of six matches, then an expected time to spend is between 4.5 hours and 6 hours for that single session.* A reasonable limitation for this research was set at sessions of six matches as the highest count, and sessions exceeding six matches will be labelled as *"Other"*. This distribution created seven unique groups to consider:

- **One**: If one match was played in the session
- **Two**: If two matches were played in the session
- **Three**: If three matches were played in the session
- **Four**: If four matches were played in the session
- **Five**: If five matches were played in the session
- **Six**: If six matches were played in the session
- **Other**: If seven or more matches were played in the session

The following is an example on how to use the session grouping: *Student A mostly played matches in sessions of three in the month with a higher win rate. At the same time, the student mostly played matches in sessions of four in the month with the lower win rate. By comparing these two, it is possible to draw the conclusion or provide the recommendation that Student A: To enact the month with a higher win rate, you should play matches in sessions of three.*

Calculating a *session distribution* can provide information about a student's frequency in terms of how often the distinct session groups occur. The maximum distribution variety a student may have is **7**, and that happens if the student has played matches in sessions of: one, two, three, four, five, six, and more than six. Example: *Student B played matches in sessions of one, two, three, and four during a month. This produces a distribution variety of **4** as the count of unique session groups is four.*

The following description is given to the distribution variety values:

- **Low**: If the count of unique session groups are <= 2, then the distribution is low.
- **Medium**: If the count of unique session groups are >= 3 AND <= 4, then the distribution is medium.
- **High**: If the count of unique session groups are >= 5, then the distribution is high.

Complementing the distribution variety, a description of the value is given (as Low, Medium, or High). Continuing the example with *Student B: As the student has a distribution variety of 4, the student has a Medium distribution variety*.

4.7 Coding Process of Qualitative Data

The dataset included unstructured text journals that was written by students. After completing a game, the students can navigate the learning management platform and reflect on their performance. The student is tasked with three different categories: 1) What did you do well? 2) What can you improve on? 3) General notes. This data entry is not mandatory, and there are no rules of subject entry. Thus, the data needs to be pre-processed and coded for further analysis. Following is a journal example taken from the dataset:

> **Positive Feedback:** Good communication and no one raged. Utility usage was splendid!
>
> **Improvement Feedback:** I shouldn't have pushed as much as I did. I played way too aggressive to surprise my enemy.
>
> **General Feedback:** I need to play less aggressive and adapt my playstyle depending on how my enemy and teammates play.

4.7.1 Data Coding

The code book was based on a grounded approach to the data, where knowledge was gained about the games, the theory in the curriculum, and courses (e.g., Mindset and Mentality, Teamwork, The History of Esports, Diversity in Esports, Item Theory, How to Work With Your Team, Team Roles). Including curriculum topics and lessons in this process was vital to gain an insight into what topics are taught and the general structure of the lessons.

Processing and working through the curriculum included content extraction and generalization of topics to identify overarching categories for journal coding. This procedure was completed for both CS:GO and LoL. There were also emerging topics that were more general (e.g., referring to external factors that are not directly tied to the game). By translating the journals (as students can freely write in English or Swedish) to English, a consistent structure was added to the data and sufficiently enabled the coding phase. The initial coding categories were: Awareness, Communication, Confidence, Early Game, Enjoyment, Late Game, Mid Game, Positioning, Progress, Self-control, Solo, and Teamwork.

The coding approach was completed at the semantic level: identifying themes and surface meanings of the data. By coding *(N=136)* journals (all from preliminary data), five themes were removed. Despite high occurrence, some themes were redundant and did not contribute to further answer the research questions. An obvious anomaly was detected, occurring only once, hence its removal. Three new topics were considered for addition to the code book, but none of them were approved as a result of redundancy and lack of relevance.

When the code book was defined, a new structured dataset was created for the coding process and validity calculation. A natural split between **Positive Feedback** and **Improvement Feedback** was then implemented, maintaining an identifiable difference between performance related to what a student did well and what a student can improve on. See Table 8 for a dataset overview.

Table 8: Journal Coding Dataset Structure

Dataset Columns	**Definition**
RowID	Row identification
Journal_ID	Journal identification
Type	Type of data
Game	Game played
User_ID	User identification
Feedback	Processed student feedback
Code X	Placeholder for code category

RowID was added for practicality and as a concise overview. **Journal_ID** is the journal identification and was used for extraction in the post-coding phase and agreement calculations. **Type** defines what data is in the dataset. In this case all entries are labeled *videogame*.

Game is linked with Journal_ID and stores a connection to the game that the feedback is associated with. **User_ID** has the student identification of the entry. **Feedback** is the column with student self-report data. **Code X** is a placeholder for any Code Book definition.

This is the part of the dataset that the coders were required to manage. The coding process was performed with a binary approach: *If the coder thinks the theme is present, then input 1. If the coder thinks the theme is absent, then input 0. None, one, some or all themes may be present in the excerpts.*

4.8 Code Book Category Definition

The combination of the curricular coding data, preliminary coding data, and a grounded approach to the definite dataset all together include the finalized code book that was used in this research. Each coding category is complemented with an example.

- **Communication** Does the student express some feedback or thoughts related to communication? This can be both in game chat, game audio or audio communication using a third party application.

  "I gave vital information to my team that produced good results in tricky situations."

- **Teamwork** Acknowledging that the game is team based and recognizing the joint effort required to play and win the game. Whenever the player mentions other teammates, teamfights or actions that include the rest of the team. This can also be related to a players performance in relation to the rest of the team.

  "I played really well this game. My positioning was not the best, but I pressured the enemies well and supported my teammates."

- **Emotional Regulation** Concerned with emotions and reactions. These entries can be wide, but some examples are: "I became really angry", or "I managed to keep calm". This is limited to a personal level, and does not include enemy reactions (e.g., "The enemy team were angry and wrote mean things in chat").

  "I kept calm for the most part, but we had communication errors and rage."

- **Game Phase** What phase of the game is the student talking about. This can range from early game, middle game, late game, Phase A, Phase B etc. It depends on the game, and it depends on the context. Based on the phase the game is in, relevant feedback may be provided for further inquiries. The different phases provide different aspects and situations in a match.

  "Played well in early game and protected my team."

- **Game Mechanics** The user is aware of an aspect they did well in, or lacked in. This can be weapon use, character use, ability use, utility use, map control, positioning, or other game defining mechanics. It is normal to see this category when concerned with feedback on a specific game objective. A total sum of all game mechanics a student has may define performance, hence its inclusion.

  "Good individual plays. I used the radar a lot to position myself well."

- **External Factors** The student talks about external factors: equipment, mood, how they slept, what they ate etc. Anything that is external in relation to the game. Based on these factors, it may seem natural to include mood as a code book category. This has not been done because students also report their mood in a Likert scale when submitting a journal after a game.

  "I should have warmed up because I was tired, therefore the start was bad."

5 Results

This chapter presents the results of the data analysis. First, results from activity data analysis and trend analysis is presented with descriptive statistics and comparisons. Second, game data is analyzed between the user groups and comparisons are made. Third, individual win rates and session analysis is presented for the top five most active students for both games, totalling ten students that were analyzed. Last, the journal coding results are presented with the most recurring codes being presented.

5.1 Activity Data Analysis

This section provides descriptive statistics and comparison of results on the activity data that originate from the learning platform. Data includes school enrollment distribution, theory activity, drill activity, exam activity, and time spent on theory. The following comparisons were made: **(1) Profile Link and Non-Profile Link**: comparison of students who applied the SRL-concept of feedback vs students who did not. **(2) Journal Entry and Non-Journal Entry**: comparison of students who applied both SRL-concepts of performance reflection and performance feedback vs students who only enacted the concept of creating performance feedback. **(3) Journal Entry and Non-Profile Link**: comparison of students who applied both SRL-concepts vs students who applied none of the concepts. Further comparisons were excluded as results would be redundant as N_{JE} and N_{NJE} are subgroups of N_{PL}.

5.1.1 School Enrollment

Table 9 presents an overview of the enrollment distribution for all students on the platform for 2020. The table shows all four schools included in the data, as well as a fifth entry for students with no school property.

Table 9: Profile Linking Rate Overview

School	Total students	Profile Linking	% Link Rate
School A	10	6	60.00
School B	28	13	46.40
School C	47	28	59.60
School D	51	9	17.60
No School Enrollment	13	7	53.85

School A had a total of 10 students, where 6 students have linked their profile, thus yielding a 60% profile link rate. School B had 28 students in total, 13 linked students, providing a 46.4% link rate. School C had 47 students, where 28 of them linked their profile, producing a 59.6% link rate. Lastly, school D had 51 students in total, with 9 linked accounts, having a total of 17.6% link rate. The 13 remaining students with no school data property included 7 students, revealing a 53.85% link rate.

5.1.2 Theory Activity

This section presents an overview of the theory activity analysis. These entries contain information about curriculum interaction on the platform. The results are presented and compared across user groups. Table 10 shows an overview of the theory activity for the groups **Profile Link** ($N_{PL=63}$), **Non-Profile Link** ($N_{NPL=86}$), **Journal Entry** ($N_{JE=23}$), and **Non-Journal Entry** ($N_{NJE=40}$).

Table 10: Theory Activity Overview

Group	Mean	Median	Lowest	Highest	Total	SD	CV
$N_{PL=63}$	96.95	77	0	370	6108	83.01	0.86
$N_{NPL=86}$	64.65	46	6	289	5560	59.58	0.92
$N_{JE=23}$	116.17	91	8	370	2672	107.16	0.92
$N_{NJE=40}$	85.90	77	0	244	3436	64.28	0.75

Profile Link and Non-Profile Link

An average student from N_{PL} read close to 97 pages of theory during enrollment, while an average student from N_{NPL} read close to 65 pages. The mean was higher for N_{PL} by 49.96%. For N_{PL} there were a total of three students with theory activity of zero. Disregard of zero activity students revealed: mean increase from 96.95 to 101.8 (5%), median increase from 77 to 85 (10.39%), and revealed the lowest entry to be 7.

A Standard Deviation (SD) of 83.01 for N_{PL} and 59.58 for N_{NPL} reflects the expected variation of theory entries per student from their respective mean. N_{PL} (0.86) and N_{NPL} (0.92) both had a Coefficient of Variation (CV) < 1, and this is considered low, meaning a low dispersion around the mean. Applying a trimmed mean of 10% was used to eliminate the influence of outliers. This calculation increased the CV for both groups. N_{PL} 0.86 increased to 0.92 and N_{NPL} 0.92 increased to 1.04. The measure of a trimmed mean of 10% did not greatly affect the central tendency of the data, revealing that the data was not heavily affected by outliers. The biggest CV increase was for N_{NPL} with an increase of 11.96%, compared to a 6.98% increase for N_{PL}.

Journal Entry and Non-Journal Entry

An average student from N_{JE} read close to 116 pages, while an average non-journal writing N_{NJE} student interacted with approximately 86 pages during enrollment. The mean was 35.24% higher for N_{JE}. In group N_{NJE} there were three students with theory activity of zero. Excluding the zero activity students revealed: mean increase from 85.90 to 92.86 (8.10%), median increase from 77.00 to 81.00 (5.19%), and revealed the lowest entry to be 7 pages for N_{NJE}. The SD values of 107.16 for N_{JE} and 64.28 for N_{NJE} shows the expected variation in theory entries per group. Both groups also had a CV < 1, which is considered low.

Calculating a trimmed mean of 10% was added to eliminate the effect caused by outliers. This revealed a CV increase for N_{JE} from 0.92 to 0.98 and an increase for N_{NJE} from 0.75 to 0.77. The trimmed mean of 10% did not greatly affect the central tendency of the data, showing that the groups are not heavily affected by outliers. The biggest CV increase was for N_{JE} with an increase of 6.52% compared to a 2.67% increase for N_{NJE}.

Journal Entry and Non-Profile Link

Comparing an average student from N_{JE} to an average student from N_{NPL} revealed a theory interaction value that was 79.69% higher for N_{JE}. Both groups had an identical CV of 0.92, which is considered low. When applying 10% trimmed mean the Non-Profile linking students had the highest CV increase.

Summary

In terms of theory interaction, the students who enacted *both* SRL-concepts (N_{JE}) had the highest average value, revealing them to be the most active students. Following, the second most active group was students who linked their accounts (N_{PL}). The least active group was N_{NPL}, revealing that students who did not apply any of the described SRL-concepts had the lowest theory interaction for 2020.

5.1.3 Drill Activity

This section gives an overview of descriptive statistics and centrality measures of the drill activity by the user groups in the dataset. Individual subsections are made to show the comparisons between specific groups.

See Table 11 for an overview of the drill activity measures for the groups **Profile Link** ($N_{PL=63}$), **Non-Profile Link** ($N_{NPL=86}$), **Journal Entry** ($N_{JE=23}$), and **Non-Journal Entry** ($N_{NJE=40}$).

Table 11: Drill Activity Overview

Group	Mean	Median	Lowest	Highest	Total	SD	CV
$N_{PL=63}$	6.94	1	0	40	437	11.05	1.59
$N_{NPL=86}$	3.23	0	0	64	278	8.8	2.72
$N_{JE=23}$	6.26	3	0	37	144	9.46	1.51
$N_{NJE=40}$	7.33	0.50	0	40	293	11.96	1.63

The data on drill activity revealed students who did not complete their drills, participation rates, and completion rates. Table 12 gives an overview of these measures for the groups **Profile Link** ($N_{PL=63}$), **Non-Profile Link** ($N_{NPL=86}$), **Journal Entry** ($N_{JE=23}$), and **Non-Journal Entry** ($N_{NJE=40}$).

Table 12: Drill Activity Participation and Completion Rates

Group	Non-Participating	% Participation Rate	Non-Completed	% Completion Rate
$N_{PL=63}$	30	52.38	21	95.19
$N_{NPL=86}$	55	36.05	40	85.61
$N_{JE=23}$	10	56.52	15	89.58
$N_{NJE=40}$	20	50.00	6	97.95

Profile Link and Non-Profile Link

The total drill activity count was higher for N_{PL} by 57.19% with entries totalling 437 compared to 278 of N_{NPL}. The mean was higher for $N_{PL}$by 114.87%. Both groups had a lowest entry at 0, meaning that both groups had students that did not complete a single drill activity. The SD of 6.94 for N_{PL} and 3.23 for N_{NPL} reveal the expected variation of drill activity entries per student in relation to their respective mean. The CV of N_{PL} and N_{NPL} were 1.59 and 2.72 respectively. CV > 1 for both groups reveal a high variance in the groups, meaning that the data is spread out around the mean.

Students with zero drill activity count up to 30 from N_{PL} and 55 from N_{NPL}. Out of all the students on the platform **Total** ($N_{Total=149}$), 58.22% did not have a single drill activity entry. The participation rate for N_{PL} was 52.38% and the participation rate for N_{NPL} was 36.05%.

N_{PL} revealed 21 non-completed entries out of the total 437 entries, which yielded a completion rate of 95.19%. N_{NPL} showed 40 non-completed entries out of the total 278 entries, resulting in a completion rate of 85.61%.

Applying a 10% trimmed mean was used to eliminate outliers (including students with zero entries). In the instance of drills, the outliers were heavily affecting the data because of the high frequency of students with zero entries. The CV of N_{PL} increased from 1.59 to 1.94 (+22.01%), and the CV of N_{NPL} increased from 2.72 to 5.09 (+87.13%).

Journal Entry and Non-Journal Entry

The total drill activity count was higher for N_{NJE} by 103.47% with entries numbering 293 compared to 144 for N_{JE}. The mean for N_{NJE} was 7.33 compared to 6.26 for N_{JE}, equaling a differential spread of 17.09%. The results show that students who did not write journals completed more drills on average. Both groups had a lowest entry at 0, meaning that both groups had students that did not complete any drill activities. The expected variation (standard deviation) of drill activity entries per student was 9.46 for N_{JE} and 11.96 for N_{NJE}. CV > 1 for both groups reveal high variance in the groups with 1.51 for N_{JE} and 1.63 for N_{NJE}.

The participation rate for N_{JE} was 56.52%, while for N_{NJE} it was 50.00%. N_{JE} had 15 non-completed drills, producing a completion rate of 89.58%. N_{NJE} had 6 non-completed drills, yielding a completion rate of 97.95%. Students who wrote journals (N_{JE}) had a higher rate of starting drills. At the same time, students who did not write journals (N_{NJE}) had a higher completion rate. To eliminate outliers, a trimmed mean of 10% was applied. The CV of N_{JE} increased from 1.51 to 1.86 (+23.17%), and the CV of N_{NJE} increased from 1.63 to 1.96 (+20.25%).

Journal Entry and Non-Profile Link

The total drill activity count for N_{NPL} with was 278 compared to 144 for N_{JE}. The average number of drill entries per student for N_{JE} was 6.26, and for N_{NPL} it was 3.23. The average student in group N_{JE} started 93.8% more drills than the average student in N_{NPL}. Both groups had a lowest entry at 0, meaning that both groups had students that did not complete a single drill activity. Standard deviation values of 9.46 for N_{JE} and 8.80 for N_{NPL} showed the expected variation of drill entries per student in relation to the mean of the group. CV values of 1.51 and 2.72 for N_{JE} and N_{NPL} respectively showed a high variance in the groups.

The participation rate for N_{JE} was 56.52%, compared to 36.05% for N_{NPL}. Calculating the completion rate of both groups resulted in 15 non-completed drills for N_{JE} and a completion rate of 89.58%. The completion rate for N_{NPL} was 85.61% with 40 non-completed drills. The results showed that students who enacted both SRL-concepts had a higher participation rate and completion rate than students who did not. Elimination of outliers in the data (trimmed mean of 10%) was applied, and the CV for N_{JE} increased from 1.51 to 1.86 (+23.17%), while the CV of N_{NPL} increased from 2.72 to 5.09 (+87.13%).

Summary

Looking at drill activity, the most active students were those that linked their profile, but did not write journals (N_{NJE}). Close behind are students that wrote journals on their performance as well (N_{JE}). The least active students in terms of drill participation and drill completion were the students that did not enact any of the SRL-principles (N_{NPL}). All groups had high CV values (CV > 1) as the count of zero-entry students were high. However, only group N_{NPL} had a CV > 2, revealing that the group were the most affected by outliers (as described by a participation rate of 36.05%).

5.1.4 Exam Activity

The exam activity section shows an overview of the exam activity and performance data for the students. Table 13 has an overview for the groups **Profile Link** ($N_{PL=63}$), **Non-Profile Link** ($N_{NPL=86}$), **Journal Entry** ($N_{JE=23}$), and **Non-Journal Entry** ($N_{NJE=40}$).

Table 13: Exam Activity Overview

Group	Exams Started	Exams Completed	Exams Graded	Exam Scores %
$\mathbf{N_{PL=63}}$	135	107	38	83.33
$\mathbf{N_{NPL=86}}$	103	75	21	82.54
$\mathbf{N_{JE=23}}$	52	41	8	91.67
$\mathbf{N_{NJE=40}}$	83	66	30	81.11

Profile Link and Non-Profile Link

Out of a total of 238 exam attempts on the platform, 135 of them were from group N_{PL} (56.72%). 103 attempts were from N_{NPL} (43.28%). N_{PL} completed 107 of the 135 exams they started, yielding a completion rate of 79.26%. N_{NPL} completed 75 out of the 103 exams they started, yielding a completion rate of 72.82%. Group N_{PL} had an average of 2.14 exams started by each student in the group, and an average of 1.7 exams completed per student. Group N_{NPL} had a marginally lower completion rate, but showed an average of 1.2 exams started per student, and an average of 0.87 exams completed in the group. An average student from N_{PL} starts 78.33% more exams than a student from N_{NPL}.

Journal Entry and Non-Journal Entry

N_{JE} completed 41 out of the 52 exams that were started, revealing a completion rate of 78.85%. N_{NJE} completed 66 out of the 83 exams the group started, thus having a 79.52% completion rate. Group N_{JE} had an average of 2.26 exams stated by each student in the group and an average of 1.78 exams completed per student. N_{NJE} had a slightly higher completion rate with 2.075 exams started per student, and an average of 1.65 exams completed. An average student from N_{JE} started 8.92% more exams than a student from N_{NJE}.

Journal Entry and Non-Profile Link

As both N_{JE} and N_{NPL} were used in previous calculations in the above sections, they were not recalculated here. Findings show that an average student from N_{JE} started 88.33% more exams than an average student from N_{NPL}.

Summary

Results of exam activity analysis revealed that the students who started the most exams and completed the most exams were students that applied both SRL-concepts (N_{JE}). The least active students (despite having the largest number of students) were students who did not apply any of the SRL-concepts (N_{NPL}). Calculating the average exam scores also showed that group N_{NPL} had the lowest score (In Table 13 it is shown that N_{NJE} had the lowest exam scores. However, as N_{NJE} is a subset of group N_{PL}, the comparison between N_{NJE} and N_{NPL} was excluded as group N_{NJE} is not representative of all students that linked their player profiles).

5.1.5 Time Spent on Theory Proxy Variable

This section gives an overview of the proxy variable for time spent on activity. All time variables are in the format of hour, minutes, and seconds (hh:mm:ss). See Chapter 4 for an overview of the proxy variable development and calculation. Table 14 provides an overview of the proxies for groups **Profile Link** ($N_{PL=63}$), **Non-Profile Link** ($N_{NPL=86}$), **Journal Entry** ($N_{JE=23}$), and **Non-Journal Entry** ($N_{NJE=40}$).

Table 14: Theory Activity Proxy Measures

Group	Relevant entries	# of blocks	Total time	Avg time spent	# Entries > 5min
$N_{PL=63}$	5946	646	36:50:10	00:00:22	162
$N_{NPL=86}$	5298	695	51:50:27	00:00:35	262
$N_{JE=23}$	2596	280	15:28:16	00:00:21	76
$N_{NJE=40}$	3350	366	21:21:54	00:00:23	86

Column *Relevant entries* was a calculation made by subtracting the number of entries exceeding five minutes from the total number of entries (Total number of entries can be seen in Table 10). The entries were excluded from the count of theory entries as the time spent in sessions that span more than five minutes are not included in the calculation that represents a time spent proxy, thus to gain the most accurate proxies for time spent, the count was also reduced. See Table 15 for an overview of the central tendency of total time spent proxies.

Table 15: Total Time Spent Proxy Central Tendency

Group	Lowest	Highest	Mean	Median	SD	CV
$N_{PL=63}$	00:00:00	02:00:26	00:35:05	00:25:08	00:29:09	0.83
$N_{NPL=86}$	00:01:09	03:39:57	00:36:10	00:23:25	00:37:57	1.05
$N_{JE=23}$	00:03:02	01:54:42	00:40:22	00:35:37	00:30:43	0.76
$N_{NJE=40}$	00:00:00	02:00:26	00:32:03	00:22:35	00:28:10	0.88

Profile Link and Non-Profile Link

The number of unique study blocks for N_{PL} was 646, and N_{NPL} it was 695. Group N_{PL} yielded an average of 10.3 unique study blocks per student, compared to 8.1 unique blocks for group N_{NPL}. The proxy for total time spent for N_{PL} equals 36 hours, 50 minutes, and 10 seconds with an average time spent per page of 22 seconds. N_{NPL} showed a total time spent of 51 hours, 50 minutes, and 27 seconds with an average time per page of 35 seconds.

N_{NPL} had both the lowest and the highest total time spent entries between the two groups. The proxy variable for average total time spent reading during enrollment was calculated relatively close between the two groups, differing only by one minute and five seconds. N_{PL} has a coefficient of variation (CV) 0.83, which is lower than 1. N_{NPL} has a CV of 1.05 which is higher than 1. This makes sense, considering N_{NPL} also revealed to include the lowest and highest proxy values of total time spent for any student, strengthening the notion that the data is more affected by outliers.

Journal Entry and Non-Journal Entry

Group N_{JE} had a total number of unique study blocks of 280, while group N_{NJE} had 366. The average number of unique study blocks per student equals 12.17 for N_{JE} and 9.15 for N_{NJE}. The proxy for total time spent for N_{JE} was 15 hours, 28 minutes, and 16 seconds, with an average time per page of 21 seconds. For group N_{NJE} the proxy for total time spent was 21 hours, 21 minutes, and 54 seconds, and an average time per page of 23 seconds.

N_{NJE} had the highest and the lowest total time spent values between these two groups. The proxy for average total time spent equals 00:40:22 for N_{JE} and 00:32:03 for N_{NJE}, revealing an average total time that was 25.95% higher for N_{JE}. Even though the journal writing group had a shorter average time spent per page, the group still had a higher total time spent average. The CV values for both groups had a value < 1, revealing a low variance in the data. However, N_{NJE} still had the highest CV value while also scoring the highest and lowest total time spent values in the dataset between the two groups.

Journal Entry and Non-Profile Link

All calculations and descriptions for groups N_{JE} and N_{NPL} can be found in the preceding sections. The average time spent per page was 66.67% higher for N_{NPL} than N_{JE}. Comparison of average total time spent values showed N_{JE} spending more time in total, while N_{NJE} spent more time on average per page.

Summary

The only group with a CV value > 1 is group N_{NPL}. On average, the students with the most unique study blocks were from group N_{JE}. The journal writing group also has the lowest time spent on average per page, while also having the highest average time spent on theory in total throughout course enrollment. Group N_{NPL} spent the most time on average per page, had the lowest number of unique study blocks, and had the highest number of entries spanning five minutes (4.71%). For all the other groups, all entries spanning five minutes made up less than 3% of all entries in the dataset.

5.2 Trend Analysis

The following section presents visualization results and trends for activity data on the platform. Visualized user groups include **Total**, **Profile Link**, **Non-Profile Link**, **Journal Entry**, and **Non-Journal Entry.**

5.2.1 Theory Activity Trend

Figure 10 is a visualization of the theory activity trend for all students on the platform ($N_{Total=149}$).

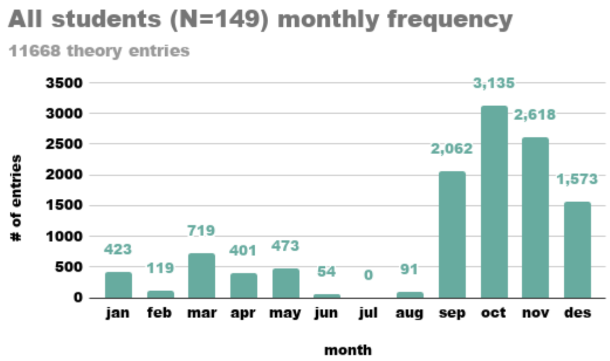

Figure 10: Theory Activity Frequency Trend For All Students

Figure 11 is a visualization of the theory activity trend for the profile linking students on the platform ($N_{PL=63}$).

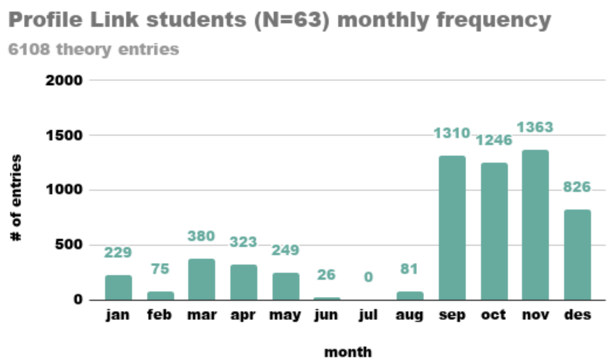

Figure 11: Theory Activity Frequency Trend For N_{PL}

Figure 12 is a visualization of the theory activity trend for the non-profile linking students on the platform ($N_{NPL=86}$).

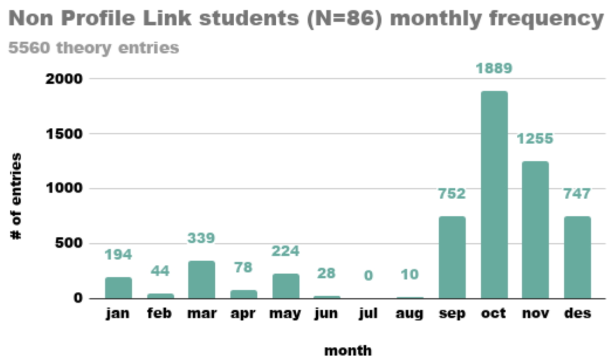

Figure 12: Theory Activity Frequency Trend For N_{NPL}

Figure 13 is a visualization of the theory activity trend for the students who linked their account to the platform and who wrote journals on their performance ($N_{JE=23}$).

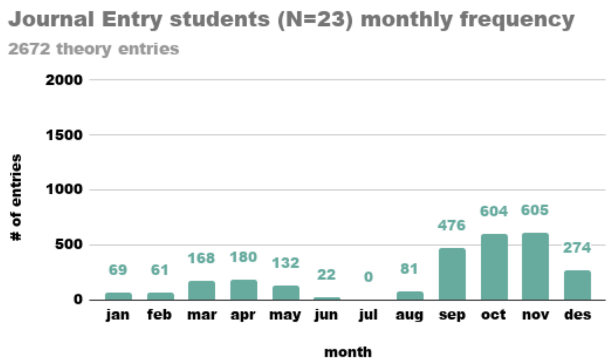

Figure 13: Theory Activity Frequency Trend For N_{JE}

Figure 14 is a visualization of the theory activity trend for the students who linked their account to the platform but did not write journals on their performance ($N_{NJE=40}$).

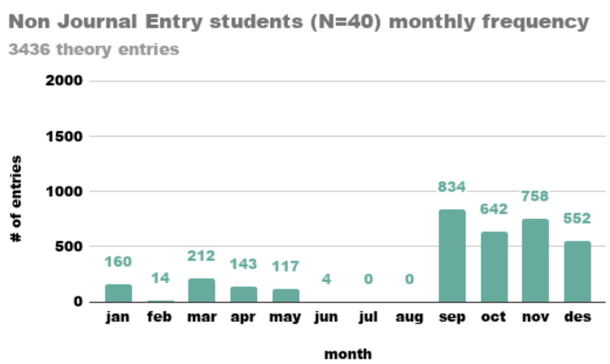

Figure 14: Theory Activity Frequency Trend For N_{NJE}

Summary

The graph visualizations for theory activity revealed a split in the data (June, July, and August). In the month of July there was not a single theory entry, and the preceding and succeeding months also revealed low counts, probably caused by holidays. All the groups followed a similar pattern of a slow start to the year, followed by growing participation, break in the summer, and a huge influx after the month of August. Despite being the smallest group, N_{JE} had 81 of 91 entries for the month of August. Comparing N_{PL} and N_{NPL} showed that the **Profile Linking** students were more consistent in their approach with less fluctuations between the months. The same patterns can be seen between N_{JE} and N_{NJE} where the **Journal Writing** students seem to be more consistent, albeit with less variation (might be caused by a smaller population sample).

5.2.2 Drill Activity Trend

Figure 15 is a visualization of the drill activity trend for all the students on the platform ($N_{Total=149}$).

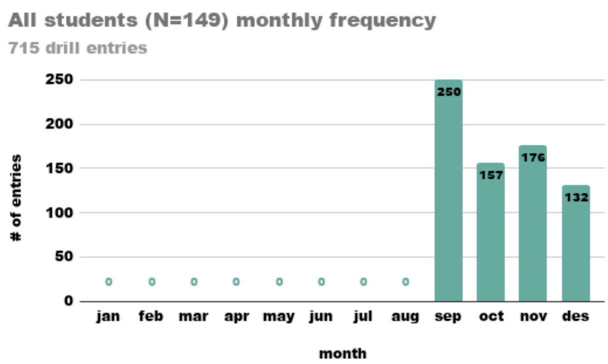

Figure 15: Drill Activity Frequency Trend For All Students

Figure 16 is a visualization of the drill activity trend for the profile linking students on the platform ($N_{PL=63}$).

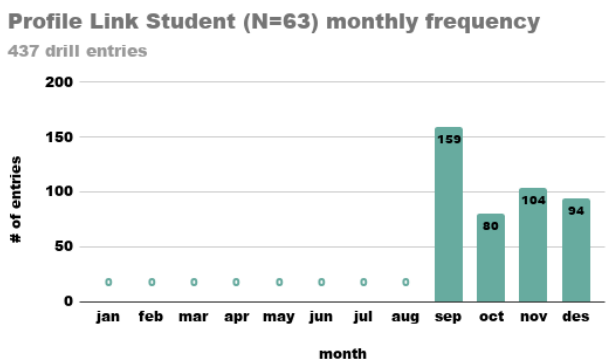

Figure 16: Drill Activity Frequency Trend For N_{PL}

Figure 17 is a visualization of the drill activity trend for the non-profile linking students on the platform $N_{NPL=86}$).

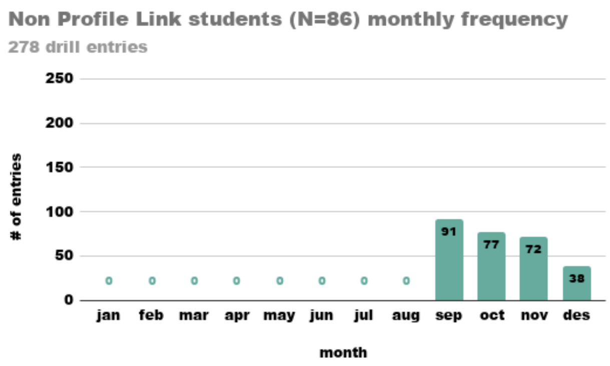

Figure 17: Drill Activity Frequency Trend For N_{NPL}

Figure 18 is a visualization of the drill activity trend for the students who linked their player account to the platform and who wrote journal entries on their performance ($N_{JE=23}$).

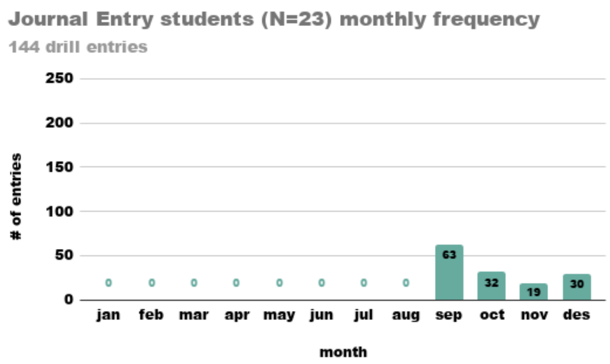

Figure 18: Drill Activity Frequency Trend For N_{JE}

Figure 19 is a visualization of the drill activity trend for the students who linked their accounts to the platform but not write any journal entries ($N_{NJE=40}$)

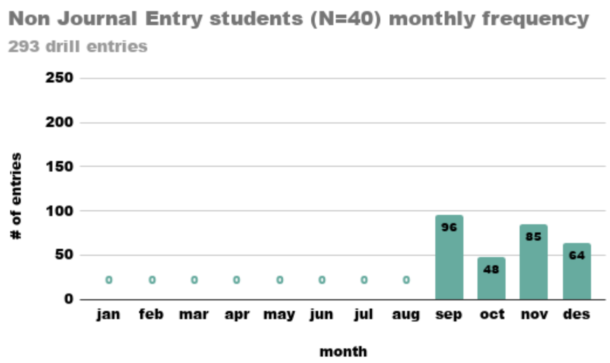

Figure 19: Drill Activity Frequency Trend For N_{NJE}

Summary

All drills that were stored in the data were timestamped in September or later in the year. By looking at the visualization for all of the students ($N_{Total=149}$) there was a decreasing trend where many drills were completed in the first month. This development can be seen in the other subgroups of students too. Comparing N_{PL} (6.94 drills per student) and N_{NPL} (3.23 drills per student) revealed that profile linking students completed a lot more drills than students who did not link their profile. N_{PL} also had an increase in the month of November in drill completions, while N_{NPL} had continuous decrease every month. Comparing N_{JE} with N_{NJE} showed that students who did not write journal entries completed more drills, and variations in drill activity could be seen in both groups.

5.3 Performance Data Analysis

Performance analysis was an analysis for in-game performance, and was thus limited to all students who opted to link their profiles to the platform **Profile Link** ($N_{PL=63}$). There were two games to consider (CS:GO and LoL), and students were distinguished based on their user group. The following sections are comparisons between **Journal Entry** ($N_{JE=23}$) and **Non-Journal Entry** ($N_{NJE=40}$).

5.3.1 Group Comparison: Counter Strike: Global Offensive

The performance analysis of Counter Strike: Global Offensive was between students who have written journal entries ($N_{JE=23}$) to the students who had linked their player profiles, but had not written any journal entries ($N_{NJE=40}$).

CS:GO: Group ($N_{JE=23}$)

Group N_{JE} had a total of 6 unique users who played CS:GO while enrolling in the course, and they made up 26.1% of the group. See Table 16 for an overview of the game performance for students in this group.

Table 16: CS:GO Matches Performance Measures N_{JE}

Student ID	# Matches	% Matches	Total Matches	% E.rate	% Win rate
269	30	3.09	30	100.00	43.33
321	583	60.10	1136	51.32	51.11
450	91	9.38	119	76.47	59.34
494	5	0.52	6	83.30	60.00
611	24	2.47	25	96.00	37.50
676	237	24.43	340	69.70	53.59
Total OR **AVG**	970	100	1656	**79.47**	**50.812**

The six students have a total of 970 games played for the duration of the year 2020. The percentage distribution of games played reveal a high variety in play-time, and two students made up 84.53% (ID 321 and 676) of all the games played in this group. By looking at the aggregated account information (i.e., data that span longer than the esports course) revealed that the majority of total games (1656) were played during enrollment in the course. The average experience rate (% E. rate) is 79.47%. A single student (ID 321) had a high frequency of games, measuring 60.1% of all the games that were played for this group. Due to this highly weighted user, the average enrollment rate was **not** calculated by

$$\frac{\sum EnrollmentGames}{\sum TotalMatchesAllPlayers} \times 100 = 58.57\% \tag{3}$$

where *EnrollmentGames* is the number of matches played during enrollment for the group, and *TotalMatchesAllPlayers* is the total number of matches played per player profile for the group. This method was used to calculate the individual experience rate. With the effect of outliers and highly weighted users, using the same approach for group average values of an experience rate would yield outcomes that do not accurately reflect a realistic result. Instead, to calculate the experience rate without being directly affected by the number of games the students have *previously* played, the following formula was used

$$\frac{\sum EnrollmentRates}{\sum Students} = 79.47\% \tag{4}$$

where *EnrollmentRates* were the individual experience rates per student in the group, and *Student* was the number of students in the group.

The majority of students in this group were new players or new profiles based on the high experience values (66.67% have an experience rate of 70% or higher, and 100% of the students show an experience rate higher than 50%). As a reminder, a lower experience rate means that a student has more experience prior to the course enrollment. Thus, a value of 79.47% indicates that these students were very new to the game. All students in this group has played the majority of their matches during enrollment. The win rate of the group provided a high variance of values, with the highest win rate measuring 59.34%, and the lowest being 37.5%, presenting a spread of 21.84%. The average win rate within the group is 50.812% for 2020, and the measure was calculated with the same approach as formula (3) and (4) for a more representative measure. **Formula 3 and Formula 4 were also used in the following sections**.

CS:GO: Group ($N_{NJE=40}$)

Group N_{NJE} had a total of 23 unique users who played CS:GO during the esports courses. These students made up 57.5% of the group. See Table 17 for an overview of these students. This group had a total of twelve students with fifteen matches or less, revealing a high variety in the number of matches played. The top four students, in terms of number of matches played, made up 76.19% of all matches. There was an average experience rate of 55.44% for this group, with experience rates varying from 1.14% to 100%. The groups' average win rate was 49.05% based on the matches played during 2020.

Less than half of the students (10 students, or 43.48%) had an experience rate lower than 50%. Considering the students who were either new or with new accounts and has played less than 15 games (12 students, or 52.17%), group N_{NJE} had more players with prior experience, as is also reflected in the lower average experience rate for the group. The win rate had a minimum value of 0%, and a maximum value of 100%.

Table 17: CS:GO Matches Performance Measures N_{NJE}

Student ID	# Matches	% Matches	Total Matches	% E.rate	% Win rate
204	171	7.32	316	54.11	56.73
213	488	20.90	1273	38.33	55.33
223	591	25.31	1542	38.33	58.71
235	65	2.78	145	44.83	58.46
238	26	1.11	41	63.41	61.54
249	455	19.06	793	56.12	49.89
255	255	10.92	565	45.13	49.80
257	2	0.09	2	100.00	50.00
455	61	2.61	103	59.22	54.10
500	5	0.21	28	17.86	60.00
520	2	0.09	2	100.00	100.00
620	1	0.04	23	4.35	0.00
625	2	0.09	7	28.57	50.00
626	6	0.26	8	75.00	66.67
627	117	5.01	130	90.00	58.97
629	11	0.47	11	100.00	54.55
632	1	0.04	1	100.00	0.00
633	2	0.09	2	100.00	0.00
634	20	0.86	195	10.26	55.00
635	14	0.60	52	26.92	64.29
680	12	0.51	1054	1.14	25.00
725	25	1.07	36	69.44	76.00
726	113	0.56	25	52.00	23.08
Total OR **AVG**	2335	100	6354	**55.44**	**49.05**

Summary

Group N_{JE} had 79.47% of their games played during the course, and Group N_{NJE} had 55.44% of their total games played during the course in 2020. The students who wrote journals had more than half of their games played during enrollment.

Results indicate that students from group N_{NJE} has played more than their counterparts prior to the course, with an average experience rate of 55.44% vs 79.47% of group N_{JE}. There were a total of 17 (or 283.3%) more students in group N_{NJE} than N_{JE}. N_{JE} had a total of 1656 games played and N_{NJE} had a total of 6354 games played. N_{NJE} had 283.7% more games played than N_{JE}. The high correlation of growth rate also revealed that the average student from each group has played 276 games each (276 for N_{JE} and 276.3 for N_{NJE}). The average amount of games per student was the same for both groups, but there was a higher win rate and a higher experience rate for N_{JE}.

5.3.2 Group Comparison: League of Legends

The performance analysis for League of Legends are between students who have written journals ($N_{JE=23}$) to the students who omitted the use of journals ($N_{NJE=40}$).

LoL: Group ($N_{JE=23}$)

Group N_{JE} had a total of 18 unique users who played LoL while enrolling in the course. These students made up 78.26% of the group. User ID 321 was present in both CS:GO and LoL calculations as the user had played both games. See Table 18 for an overview of the students in this group.

Table 18: LoL Games Performance Measures N_{JE}

Student ID	# Games	% Games	Total Games	% E.rate	% Win rate
31	502	5.05	1876	26.76	47.81
33	632	6.36	2271	27.83	46.52
46	412	4.14	942	43.74	51.70
100	113	1.14	1004	11.25	48.67
233	1178	11.85	3110	37.88	51.53
321	75	0.75	242	30.99	60.00
452	33	0.33	175	18.86	51.52
453	359	3.61	733	48.98	55.15
489	873	8.78	3751	23.27	50.29
490	631	6.35	1765	35.75	53.57
491	1790	18.01	5529	32.37	52.40
492	884	8.89	2313	38.22	52.38
493	994	10.00	2765	35.95	53.72
503	603	6.07	1436	41.99	56.88
505	49	0.49	49	100.00	51.02
508	62	0.62	63	98.41	51.61
509	510	5.13	1099	46.41	50.20
640	241	2.42	1492	16.15	53.11
Total OR **AVG**	9941	100	30615	**39.71**	**52.12**

The 18 users have a total of 9941 games played during enrollment. The percentage weight of all the games played reveal a somewhat scattered distribution where certain users are very active, others are somewhat active, and some are inactive. The most active students (students with games > 800) make up 57.53% of the games (ID 233, 489, 491, 492, and 493). Comparing matches played during enrollment to the student's player profiles (which account for data before the course) exposes an average experience rate (% E. rate) of 39.71%. The win rate has a moderate variance, with the lowest value being 46.52%, and the highest value showing 60.00%, producing a spread of 13.48%. The average win rate for group N_{JE} equals 52.12%.

LoL: Group ($N_{NJE=40}$)

Group N_{NJE} has a total of 20 unique students who enrolled to the course and played LoL. The 20 students make up 50.00% of the group. See Table 19 for an overview of the students in this group.

Summarizing the amount of games played by the 20 students during enrollment resulted in 10960 games played. The percentage weight of games played were somewhat scattered, with highly active, moderately active and low activity between users. The students with highest activity (students with games > 800) made up 54.67% of the matches played (ID 220, 639, 677, and 828).

Table 19: LoL Games Performance Measures N_{NJE}

Student ID	# Games	% Games	Total Games	% E.rate	% Win rate
93	91	0.83	293	31.06	58.24
101	174	1.59	425	40.94	47.70
207	34	0.31	293	11.60	52.94
210	19	0.17	151	12.58	47.37
211	378	3.45	1120	33.75	51.85
216	408	3.72	1222	33.39	53.43
220	1222	11.15	1696	72.05	49.92
221	340	3.10	901	37.74	51.47
223	70	0.64	652	10.74	44.29
242	677	6.18	2326	29.11	48.89
520	191	1.74	191	100.00	43.98
620	311	2.84	656	47.41	50.80
637	101	0.92	241	41.91	50.50
638	193	1.76	359	53.76	48.70
639	1784	16.28	4018	44.40	50.34
641	440	4.01	1663	26.46	51.14
677	1865	17.02	4268	43.70	51.58
679	787	7.18	1682	46.79	49.05
686	755	6.89	1791	42.16	52.19
828	1120	10.22	1494	74.97	49.64
Total OR **AVG**	10960	100	25442	**41.72**	**50.20**

The average experience rate (% E. rate) equals 41.72%, revealing that the average student played 41.72% of their games during enrollment. There was a moderate variance in the win rate for the group. The lowest value was 43.98%, and the highest value was 58.24%, resulting in a 14.26% spread. The average win rate for group N_{NJE} equals 50.20.

Summary

None of the groups showed a higher variance than +/- 15% for the win rate, providing a moderate spread for both of the groups. As introduced in Section 5.3.1, the analyses measuring average values follow Equation 4. Group N_{JE} had 27.78% of their total games played during the course, while group N_{NJE} had 35.38% of the matches completed in enrollment. Students who have written journals had less of their games played in enrollment than users who did not write journals. The average experience rate was lower for N_{JE} (39.71%) than N_{NJE} (41.72%), hence the journal writing students had played more prior to enrollment. There were a total of 2 (or 11.11%) more users for N_{NJE} than N_{JE}. Total matches for journal writing students were 35780, compared to 30974 for the non-journal writing students. The group with more people had played less games in total. The average number of matches for N_{JE} was 1987.78 matches per student, and N_{NJE} revealed an average of 1548.70 games per student. There has been played, on average, more matches per student in the journal writing group, while also having a lower experience rate, and a higher win rate.

5.4 Individual Performance and Session Analysis

Counter Strike: Global Offensive

Individual performance analysis for CS:GO included the top five most active students through 2020. The measure of activity is counted as the number of matches played through the course. Included students are (ID 213, 223, 249, 255, and 321). All students are part of the profile linking group ($N_{PL=63}$).

5.4.1 Student 213

Student 213 is part of the non-journal writing group ($N_{NJE=40}$). The student had a total of 488 played matches, a win rate during the course of 55.33%, and at least one game entry per month of 2020. This student had an experience rate of 38.3%, meaning that the 488 games played during 2020 equals 38.3% of all the games played for this account. The analyzed months were April (higher deviation), October (lower deviation), and December (most played matches).

In April there were 58 games played with 38 wins and 18 unique dates. The win rate of 65.52% is a point increase of 10.19% compared to the mean of the course. The games were played at daytime and early evening.

The monthly trend was distinguished by sessions that comprised mainly one or two games (55.17% of games are played in sessions of one or two). There was high variety in session distribution (6). An average of 3.2 games was played per unique date.

In October there were 40 games played with 18 wins and 15 unique dates. The win rate of 45% is a point decrease of -10.33% compared to the mean of the course. Most games were played at daytime and in the evening. The most played sessions were one game sessions (27.50%) and three game sessions (52.50%). There was medium variety in session distribution (4). There was an average of 2.7 games player per unique date.

The month with most games was December. There were 75 games played and 19 unique dates. Timestamps of this month correlated well with April and October, apart from some entries early in the morning. The win rate for this month was 49.33%. December revealed a mixture of results from the deviating months. Three-game sessions were dominant, but there was a high variety (6) in session distribution, while also seeing sessions with > 4 games more frequently. Less than half of the wins came after a previous win. There was an average of 3.9 games played per unique date.

Student 213 had a higher win rate when the most recurring session types consisted of one or two matches. A lower win rate was shown when the most recurring sessions consisted of one or three matches. Looking at the most active month, three matches were the most common for the sessions. Keeping sessions to one or two matches, close to an average of three matches per day, and high variety in sessions will reflect the tendencies seen in the month with the highest win rate.

5.4.2 Student 223

Student 223 was part of the non-journal writing group ($N_{NJE=40}$). The student has played 591 matches in the course with a win rate of 58.71%. The student had at least once game entry per month of 2020.

The 591 matches make up an experience rate of 38.3%. The analyzed months were November (higher deviation), May (lower deviation), and June (most played games).

In November there were 18 games played with 13 wins and 11 unique dates. The win rate was 72.22%, a point increase of 13.51% from the mean. The majority of games were played in the evening. There was medium distribution in session types (3). The games were played in sessions of either one, two or three games. Two-game sessions were prominent, representing 55.56% of the matches played. There was an average of 1.6 games played per unique date.

May showed 45 matches, 20 wins and 13 unique dates. The win rate was 44.44%, a decrease of -14.27% considering the course rate. Most games were played in the evening. There was high distribution in session categories (6).

Sessions of three games were the most played (26.67%), followed by sessions of two (17.78%), and sessions of four (17.78%). An average of 3.5 games were played per unique date.

The month with the highest match count was June. There were 219 games over 28 unique dates, and a win rate of 62.10%. Timestamps did not match well with the deviating months as matches in June were played at all times of the day. The high count of games and unique dates revealed high activity, with a substantially high distribution (more than 6) of session counts. 43.38% of the games were played in sessions of seven or more games. Excluding sessions of seven and higher, the most recurring sessions were games of five (18.26%) and equally many of four and six (10.96% each). For June, there was an average of 8 games per unique date.

Student 223 had a higher win rate when the majority of sessions were limited to two matches. The lowest win rate was in a month with the most recurring sessions consisting of three matches, and the most active month had mostly sessions of seven or more matches. For this student, sessions of two matches, an average of two matches per day, and a high variety in sessions will emulate the month with the highest win rate.

5.4.3 Student 249

Student 249 was part of the non-journal writing group ($N_{NJE=40}$). This student has played 445 matches, and had a win rate of 49.89%. For each month of 2020, at least one game has been played. The analyzed months were May (higher deviation), September (lower deviation), and October (most games played).

In May there were 56 games, 35 wins and 16 unique dates. The win rate was 62.50%, an increase of 12.61% compared to the rest of the year. Larger parts of the matches were played during daytime, with only a few in the morning and evening. The count of unique session groups revealed a high distribution (5). Most games were played in session of three (48.21%), followed by sessions of two (25%). An average of 4.3 matches were played per unique date.

For September the student had 53 games, 19 wins and 19 unique dates with a win rate of 35.85%. The win rate has a decrease of -14.04% from the mean. Most of the games were played during daytime. There was high distribution variety in session groups (5). Most games were played in sessions of two (52.83%), followed by sessions of five (18.87%). An average of 2.8 games were played per unique date.

The month with highest match count was October with 63 games, 27 wins and 20 unique dates. The win rate is 42.86%, a decrease of -7.03% compared to the mean. Timestamps of the games followed the same patterns as May and September. Session variety distribution was medium (3). Games were either played in sessions of three (42.86%), sessions of two (34.92%), or sessions of one (22.22%). An average of 3.15 games were played per unique date.

Student 249 had a higher win rate when the majority of sessions comprised of three matches. The month with the lowest win rate had sessions of two or five matches as the most common session types. The most active month also saw three matches as the most recurring session type. For student 249, having sessions of three matches, an average close to four matches per day, and a high variety in session distribution will emulate the tendencies of the most winning month.

5.4.4 Student 255

Student 255 was part of the non-journal writing group ($N_{NJE=40}$). Through enrollment, there had been played 255 games with a win rate of 49.80%. At least one game was played each month of 2020. The analyzed months were January (higher deviation), September (lower deviation), and March (most played games). January saw 27 games, 13 unique dates and a win rate of 59.26%. The win rate had an increase of 9.46%. The matches were played at morning or daytime. There was medium distribution variety in session groups (4). Most games were played in sessions of two (66.67%). An average of 2.1 games were played per unique date.

In September there were 15 games played, with 3 wins and 10 unique dates. A win rate of 20% is a -29.8% decrease from the monthly average win rate. The games were played at daytime or in the evening. The variety of session group distribution was medium (3). Most games were played in sessions of one (55.33%). The average game per unique date was 1.5.

March was the month with the most matches played. There were 38 games, with 19 wins, 17 unique dates, and a win rate of 50%. That was an increase of 0.2% from the win rate mean. The timestamps of matches were mainly at daytime, but also present at morning and evening, seeing some difference from January and September. Session variety distribution was medium (4). The majority of the matches were played in sessions of two (31.58%) and three (47.37%). An average of 2.2 games were played per unique date.

Student 255 had the highest monthly win rate when sessions consisted mainly of two matches, compared to a majority of sessions being comprised of one match in the lower deviating month. The most active month revealed three and two matches to be the most common. For this student, playing sessions of two matches, maintaining an average of two matches per day, and applying a medium variety of sessions will emulate the highest winning month.

5.4.5 Student 321

Student 321 was part of the journal writing group ($N_{JE=23}$). The player had a total of 583 matches with a win rate of 51.11%. There was at least one games per month of 2020. The analyzed months were September (higher deviation), May (lower deviation), and June (most played matches).

September had 27 games, 13 unique dates and a win rate of 59.26%. The win rate increase was 8.15% compared to the mean. All the matches were played in the evening. The variety of session group distribution was medium (4). The majority of the games were played in sessions of three (44.44%). The average game per unique date was 2.1.

In May, the student played 18 matches, with 8 wins and 8 unique dates. The win rate of 44.44% was a decrease of -6.67% from the mean. All matches were played in the evening or at night. There was medium variety in session distribution (4). A majority of the games were played in either sessions of one (33.33%) or in sessions of three (33.33%). The average game per unique date was 2.25.

The month with the most matches was June with 100 matches, 54 wins, and 24 unique dates. The win rate of 54% was an increase of 2.89% from the mean. Matches were played either in the evening or at night, much the same as May and September. A high session variety distribution was established (number here), with 23% of the matches being played in sessions of seven or more games. Apart from that, the most common session groups were sessions of four games (28%) and sessions of six games (18%). There was an average of 4.2 games played per unique date.

Student 321 saw a higher win rate when most sessions consisted of three games. The lowest month had as many sessions of one as of three, but the most active month had seven or more matches as the most recurring session length. To reflect the trends of the most winning month, this student can: keep sessions to three matches, omit one-game sessions, play an average of two games per day, and keep session variety to a medium distribution.

Summary

Between the top five most active CS:GO students, there were 186 games played in total in their months of higher deviation (See Table 20), and 171 matches played in total in their months of lower deviation (See Table 21). The table shows the distribution overview per student, the total number of sessions, the number of games, and the percentage weight of each session grouping.

Table 20: CS:GO Higher Deviating Session Overview

Matches in session	**One**	**Two**	**Three**	**Four**	**Five**	**Six**
Student 213	14	9	2	1	2	1
Student 223	5	5	1	0	0	0
Student 249	6	7	9	1	1	0
Student 255	2	9	1	1	0	0
Student 321	5	3	4	1	0	0
Total	32	33	17	4	3	1
Number of matches	32	66	51	16	15	6
% of matches played	17.20	35.48	27.42	8.60	8.07	3.23

Table 21: CS:GO Lower Deviating Session Overview

Matches in session	One	Two	Three	Four	Five	Six
Student 213	11	2	7	1	0	0
Student 223	2	6	4	2	1	1
Student 249	5	14	2	1	2	0
Student 255	8	2	1	0	0	0
Student 321	6	1	2	1	0	0
Total	32	25	16	5	3	1
Number of matches	32	50	48	20	15	6
% of matches played	18.71	29.24	28.07	11.70	8.77	3.51

League of Legends

Individual performance analysis for LoL included the top five most active students through the course in 2020. Activity was defined as the number of games played through the course. Included students were (ID 220, 233, 491, 639 and 677). All included students were part of the profile linking group ($N_{PL=63}$).

5.4.6 Student 220

Student 220 was part of the non-journal writing group ($N_{NJE=40}$). The student played 1222 matches with a win rate of 49.92%. The analyzed months were September (higher deviation), August (lower deviation), and May (most played matches).

In September there were 160 games, 26 unique dates, and a win rate of 51.25%. The win rate was an increase of +1.33% compared to the win rate mean in the course. There was some variety in the timestamps of the games, but the majority were played in the evening or at night. There was high distribution variety in session groups (7). Most of the games were played in sessions of seven or more (33.13%), followed by session of six games (18.75%). The average game per unique date was 6.15.

For August, the student played 61 matches, had 20 unique dates, and a win rate of 42.62%. Compared to the average win rate in the course, August had a decrease of -7.3%. Most games were played in the evening or at night. There was high distribution variety in session groups (5). The majority of the matches were played in sessions of two (26.23%) or four (26.23%). An average of 3.05 games were played per unique date.

The month with the most matches was May with 299 games, 28 unique dates and a win rate of 50.50%. The win rate was an increase from the mean of 0.58%. Matches were played in the morning, at daytime, in the evening and at night, with most of them being late in the evening or at night (more distributed than September and August).

Session variety distribution was high (7), and the most common sessions groups were sessions with more than seven games (63.21%), and subsequently the most common sessions were games of four (10.70%) and three (9.03%). The average number of games played per unique date was 10.68.

Student 220 had the highest monthly win rate consist mainly of sessions with seven matches or more. The lowest month mainly consisted of sessions of two or four matches, and the most active month had the majority of matches played in sessions of seven or more. For student 220 to enact the highest winning month, then sessions should mostly consist of seven or more matches, limit two-match sessions, play an average of 6 games per day, and maintain a high variety in session distribution.

5.4.7 Student 233

Student 233 was part of the journal writing group ($N_{JE=23}$). For 2020, the student has played 1178 matches with a win rate of 51.53%. The analyzed months were March (higher deviation), December (lower deviation), and October (most active month).

In March there were 171 games, 30 unique dates and a win rate of 56.14%. The win rate was an increase of 4.61% compared to the win rate mean of the course. Most of the games were played during daytime or early evening. The variety of session group distribution was high (7). Most games were played in sessions of seven games or more (25.15%), four games (18.71%) or six games (14.04%). An average of 5.7 games were played per unique date.

For December there were played 20 games with 12 unique dates and a win rate of 40%. The win rate was a decrease of -11.53% from the mean. All games were played either late afternoon or in the evening. There was low distribution variety in session groups (2). All matches played were either in sessions of one (50%) or sessions of two (50%). An average of 1.67 games was played per unique date.

The month to analyze with the most matches was October (second most active month, but March was excluded as it was flagged). For October there was 128 games, 29 unique dates and a win rate of 51.56%. The win rate was an increase from the mean of 0.03%. Games were played at all times of the day, with the majority being in the afternoon or evening. Session variety distribution was high (7), with the most common session groups being sessions of five games (27.34%), sessions of three games (21.09%), and sessions of two games (18.75%). The average number of games played per unique date was 4.41.

For the month with the highest win rate, student 233 had the majority of matches being played in sessions of seven or more. The lowest month in terms of win rate had only sessions of one or two matches, and the most active month had most matches played in sessions of five.

To enact the highest winning month, the student should: keep the majority of sessions to seven or more matches, omit one-match and two-match sessions, play an average of six matches per day, and maintain a high variety in session distribution.

5.4.8 Student 491

Student 491 was part of the journal writing group ($N_{JE=23}$). While enrolling in the course, the student has played 1790 games with a win rate of 52.40%. The analyzed months were December (higher deviation), March (lower deviation), and May (most played matches).

For December there were 57 games played, 18 unique dates and a win rate of 57.89%. The win rate was an increase of 5.49% compared to the win rate throughout the course. The majority of the games were played in the afternoon and evening, with a few in the morning or generally at daytime. Most games were played in sessions of two (24.56%), sessions of three (21.05%) or sessions of one (21.05%). The average game per unique date was 3.17.

In March, there was a total of 66 games played over 18 unique dates, and with a win rate of 45.45%. The win rate was a decrease of -6.95% compared to the win rate of the course. Match timestamps showed games in the morning, daytime, evening and night, with highest frequency at afternoon and evening. Session variety distribution was high (5), with the most common session groups being sessions of three games (45.45%), one game (21.21%) and two games (15.15%).

May was the most active month with 259 games, 29 unique dates and a win rate of 52.90%. The win rate was a deviation of +0.50% compared to the win rate throughout the course. The majority of games were played during daytime, often starting sessions before noon and playing throughout the day. Session variety distribution was high (7), and the most common sessions groups were sessions with more than seven games (66.80%). Sessions of three (9.27%) and four (7.72%) were the two most common after game sessions of >=7. The average number of games player per unique date was 8.93.

For student 491, the month with the highest win rate consisted mainly of sessions with one, two, or three matches (very close percentages between the sessions). The lowest scoring month had the majority of games played in sessions of one and three. Sessions of the most active month consisted mainly of seven or more matches. To enact the highest winning month, student 491 should mainly play sessions of two, averaging close to three games per day, and maintaining a high variety in session types.

5.4.9 Student 639

Student 639 was part of the non-journal writing group ($N_{NJE=40}$). The player had a total of 1784 matches with a win rate of 50.34%. The analyzed months were February (higher deviation), August (lower deviation), and July (most played matches).

In February there was a total of 123 matches with 23 unique dates, and a win rate of 58.54%. Compared to the win rate through the year, February had an increase of +8.2%. The majority of the games were played at daytime or early evening. There was high distribution variety in session groups (7), with the most common sessions being games of three (29.27%), games of two (19.51%), and games of six (19.51%). An average of 5.35 games were played per unique date.

In August there were 235 games, 30 unique dates and a win rate of 44.26%. Compared to the win rate of the course, August had a decrease of -6.08%, while also being the most active month. There was variety in timestamps for the games played, but there was more presence of games at early evening and also at night. The variety of session group distribution was high (7). More than half of the games were played in sessions of seven games or more (59.15%). Subsequently, sessions of five (12.77%), and three (7.66%) are the most frequent. The average number of games played per unique date was 7.83.

For July there were 214 games played, with 30 unique dates and a win rate of 54.21%. The win rate was an increase of +3.87% from the average win rate of the course. The majority of the games were played in the evening. The variety of session group distribution was high (7). The most frequent session comprise of seven or more games (48.13%), with sessions of three (12.62%), four (11.21%), and six following behind (11.21%). The average number of games played per unique date was 7.13.

For student 639, the highest winning month mainly had sessions of two and three matches. The lowest rated month had the majority of matches played in session of seven, and the same goes for the most active month. For this student to continue the trend of the highest winning month, then sessions should be limited to two or three matches, removing longer sessions of five, six, seven, or more matches, play an average of 5.35 matches per day, and maintain a high variety in sessions.

5.4.10 Student 677

Student 677 was part of the non-journal writing group ($N_{NJE=40}$). For 2020, the student was a total of 1865 games played with a win rate of 51.58%. The analyzed months were April (higher deviation), July (lower deviation), and September (most active month).

In April there were 75 games, 18 unique dates and a win rate of 64.00%. The win rate was an increase of +12.42% compared to the average win rate for 2020. Games were played at a variety of times, but the most frequent timestamps were in the afternoon and evening. There was high distribution in session types (6). Matches were played in sessions of two (24.00%), sessions of seven or more (22.67%), sessions of one (20.00%), and sessions of three (20.00%). There was an average of 4.17 games played per unique date.

For July there were 119 matches, 18 unique dates and a win rate of 41.18%. Compared to the average win rate of 2020 for this student, there was a decrease of -10.40%. The majority of games were played either in the afternoon or the evening. There was high distribution in session types (7). More than half of the games were played in sessions of seven games or more (52.10%). Next, the most common session types were two-game sessions (13.45%), and three-game sessions (10.08%). An average of 6.61 games were played per unique date.

September was the month with the highest game count at 241 games, 29 unique dates, and a win rate of 52.28%. The win rate was an increase of +0.70% compared to the rest of the year. Most games were played in the afternoon or in the evening. Session variety distribution was high (7). 41.08% of the matches were played in sessions of seven or more games. Apart from that, the most common session groups were sessions of five games (16.60%), sessions of three games (13.69%), and sessions of six games (12.45%).

For this student, the highest winning month mainly had sessions of one, two and three matches, and sometimes seeing more than that. The lowest scoring month had the majority of matches played in sessions of seven or more, and the same is also for the most active month. For student 677 to maintain the same trend from the highest scoring month, then sessions should be mainly limited to one, two, or three matches for most times, remove sessions of seven or more matches, keep an average of four matches per day, but also maintain a high variety in sessions.

Summary

For the top five most active LoL students, there were 586 matches played in their months of higher deviation (See Table 22), and 501 games played in total for their months of lower deviation (See Table 23). The tables shows the distributions per student, the total number of sessions, the number of games, and the percentage weight of each session grouping.

Table 22: LoL Higher Deviating Session Overview

Matches in session	**One**	**Two**	**Three**	**Four**	**Five**	**Six**	**Other**
Student 220	5	7	8	3	4	5	6
Student 233	14	11	7	8	3	4	5
Student 491	12	7	4	0	1	1	1
Student 639	1	12	12	3	2	4	2
Student 677	16	9	5	1	0	1	2
Total	48	46	36	15	13	15	16
Number of matches	48	92	108	60	65	90	123
% of matches played	8.19	15.70	18.43	10.24	11.09	15.36	20.99

Table 23: LoL Lower Deviating Session Overview

Matches in session	**One**	**Two**	**Three**	**Four**	**Five**	**Six**	**Other**
Student 220	10	8	4	4	0	0	1
Student 233	10	5	0	0	0	0	0
Student 491	14	5	10	0	1	0	1
Student 639	8	6	6	4	6	2	11
Student 677	5	8	4	2	2	1	5
Total	47	32	24	10	9	3	18
Number of matches	47	64	72	40	45	18	215
% of matches played	9.39	12.77	14.37	7.98	8.98	3.59	42.91

5.5 Journal Analysis

For each match that has been played, a student can voluntarily input feedback on their performance in the structure: *(1) What did you do well?, (2) What can you improve on?, and (3) General Notes.* In this instance, General Notes were excluded due to a high frequency of unrelated entries (i.e., not related to the game, invalid entries, empty entries).

In the dataset, there was a total of 81 journal entries to consider for analysis, and each journal had unique input on Positive Feedback and unique input for Improvement Feedback. In combination, there were 162 entries (connected as pairs) to analyze. Each journal address a single match that has been played. Out of the 81 journals, 38 (46.91%) are for CS:GO. LoL make up 43 (53.09%) of the journals. 20 out of the 81 entries for positive feedback were invalid (24.69%), and 14 out of the 81 entries for negative feedback were labelled invalid (17.28%). Journal entries were labelled as invalid if the entry was empty or incomprehensible. After removing invalid entries, there were 61 entries for positive feedback and 67 entries for improvement feedback, creating a total of 128 journal entries to code.

5.5.1 Coding Reliability and Results

An identical coding scheme was applied to both the Positive Feedback data and Improvement Feedback data. Three coders were involved, and each coder had individually coded the data with added guidance of the code book and coding definitions. Percentage inter-rater reliability (P_A) is calculated by:

$$P_A = \frac{N_A}{N_A + N_D} \times 100 \tag{5}$$

where N_A is equal to the total number of agreements and N_D equals the total number of disagreements. The agreement rate between the three coders is:

$$P_A = \frac{761}{761 + 135} \times 100 = 84.93. \tag{6}$$

A percentage inter-rater reliability score of 84.93% revealed a high agreement rate between the coders. For Positive Feedback and Improvement Feedback, the percentage scores were 85.95% and 84.01% respectively. As three people were involved in the coding phase, the final coding results were based on how the majority of the coders interpreted the data. If two coders input 1 for any coded category (i.e., the theme is present), the final coding results would then consider the theme to be present for that given journal, despite one coder disagreeing. See Table 24 for an overview of code category frequency for the positive feedback. See Table 25 for an overview of code category frequency for the improvement feedback.

Table 24: Positive Feedback Code Frequency

Positive Feedback (N=61)	***Comm.**	***SP**	***Team**	***ER**	***GP**	***GM**	***EF**
# Entries	11	51	17	6	20	29	0
% Frequency	18.03	83.61	27.87	9.84	32.79	47.54	0

Table 25: Improvement Feedback Code Frequency

Improvement Feedback (N=67)	***Comm.**	***SP**	***Team**	***ER**	***GP**	***GM**	***EF**
# Entries	9	55	11	12	10	36	6
% Frequency	13.43	82.09	16.42	17.91	14.93	53.73	8.96

Column **Feedback** differentiates feedback related to positive aspects of the match, negative aspects, or improvement features of the match. **Comm** is short for communication, and is related to aspects of communication within the game, both textual and audio. **SP** is an abbreviation for Solo Performance, and is related to aspects and descriptions of the student's performance. **Team** is short for Teamwork, and is used when the player acknowledges that the game is team based, mentions other teammates or actions that include the rest of the team. **ER** is short of Emotional Regulation, and is concerned with emotions and reactions. **GP** is an abbreviation for Game Phase, and identifies the phase of the game (e.g., early, late, first-half, second-half). **GM** is short for Game Mechanics and covers game related mechanics (e.g., weapon use, character use, ability use, utility use, positioning, or game specific mechanics). **EF** is short for External Factors and is related to events or causes not directly tied to the game.[18]

[18]For a full overview of the data processing, thematic analysis, and code book definition, see Chapter 4.

For journals on positive feedback, the most used code was related to Solo Performance aspects (83.61%). Game mechanics is the second most prevalent category (47.54%), and themes on Game Phases is the third most common (32.79%). The fourth prevailing category is Teamwork (27.87%), followed by Communication (18.03%). The two least common categories are Emotional Regulation (9.84%) and External Factors (0%).

For journals on improvement feedback, the most common category was Solo Performance (82.09%), followed by Game Mechanics (53.73%) and thirdly Emotional Regulation (17.91%). The fourth most common category is Teamwork (16.42%). The fifth category is Game Phase (14.93%). The two least coded categories for improvement feedback is Communication (13.43%) and External Factors (8.96%).

6 Visualizations

6.1 Esports Dashboard

Based on the available student data, Figure 20 visualizes the learner model structure for the esports management platform based on the available data. The learner model can provide data related to Performance Data and Activity Data. Performance data is part of the voluntary events that students may choose to do (i.e., connecting their game account and write journals on their performance). Activity data is automatically captured data on the learning management platform that produces a record of theory activity, drill activity, exam activity, school enrollment, and training program participation. This chapter provides dashboard prototype examples.

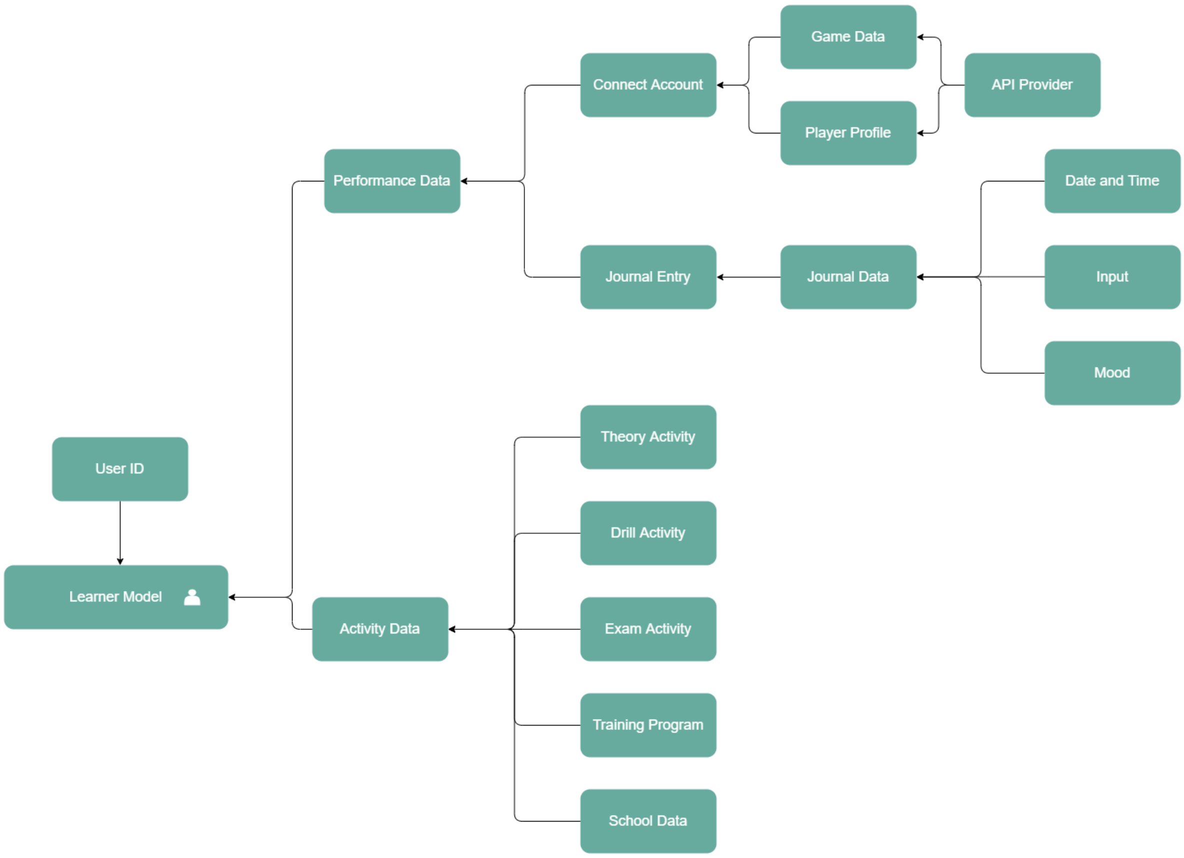

Figure 20: Learner Model Structure

6.1.1 Visualization Examples

The visualization examples are based on data availability and analysis results. Data may be presented in textual form or through visualizations to provide insightful presentations for the interpreter.

Student Landing Page

Figure 21 illustrates a prototype landing page for a student facing dashboard application. Users can access their statistics, including data on Performance, Theory, Drills, Journals, Trends and Leaderboards. By applying an editable functionality, a user may adapt their dashboard to suit their needs and provide methods that fit their learning strategies.

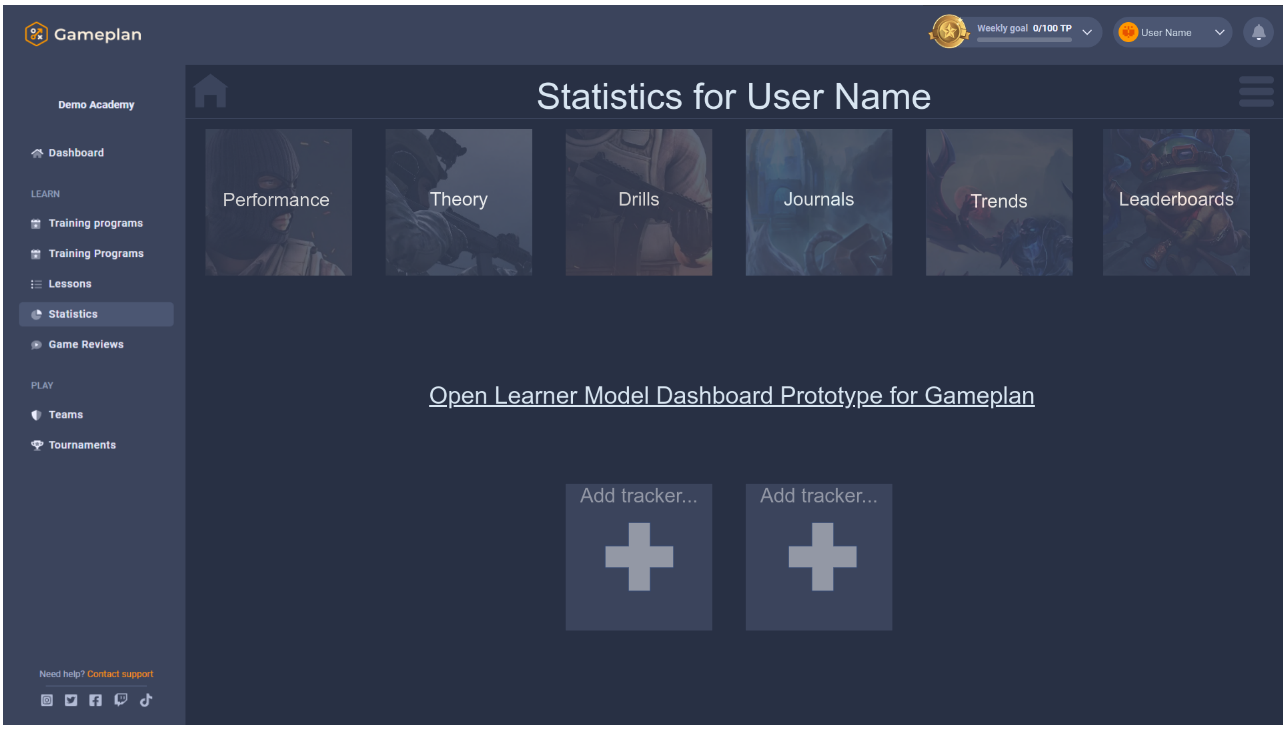

Figure 21: Prototype Image 1: Landing Page

Student Performance Overview

Figure 22 shows a proposed design for visualizations of game performance data for a student. Total aggregated player profile statistics are provided at the top of the page. Past performance provides a more detailed overview of specific statistics that is editable based on data availability. Past performance can be filtered with 'Last Month', 'Last Week', 'Custom' or 'Deviation'. By filtering on deviation, a user may look at months with a higher or lower than usual deviation from an average statistic, thus providing an opportunity to view the matches in depth to identify what went particularly well, or what went wrong.

Game frequency is a graph visualization to provide an overview of the number of matches played and match frequency. Visualization of such data can provide an overview for students on their playing behavior. The graph can be complemented with timestamps to increase its functionality. Combining performance data with timestamps may reveal hidden information otherwise not available.

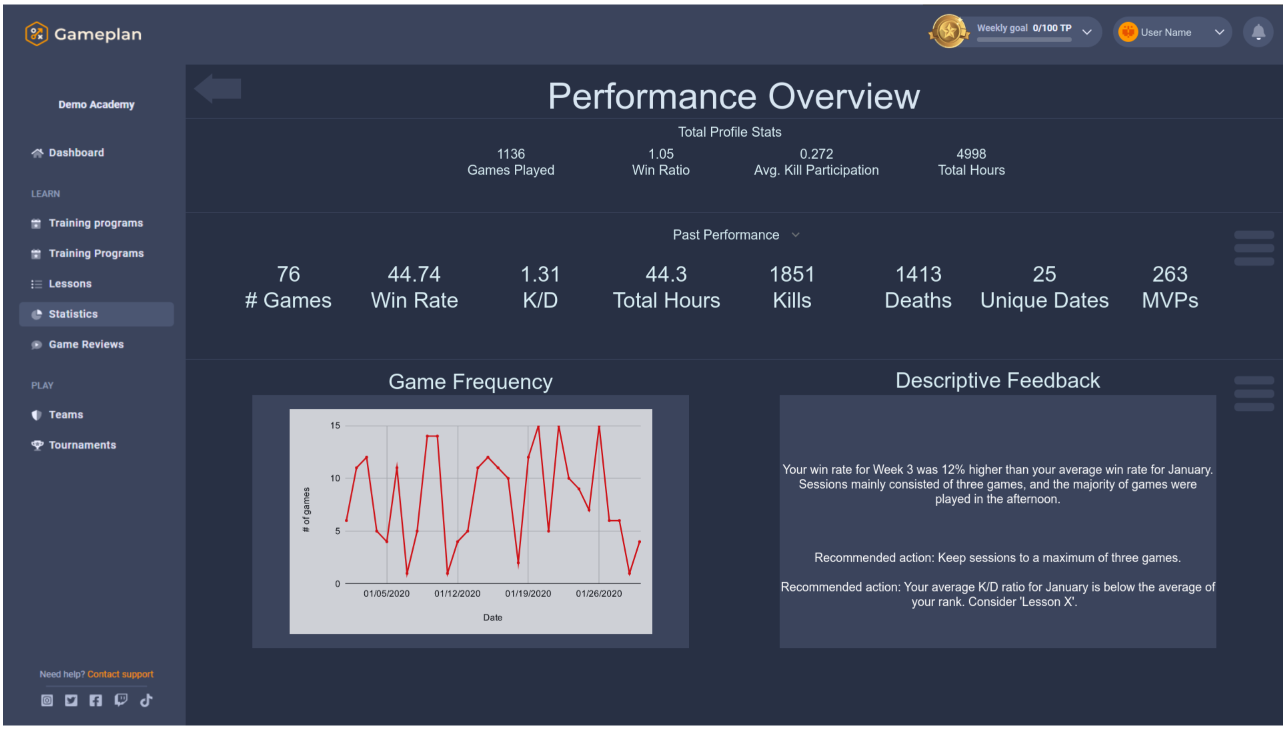

Figure 22: Prototype Image 2: Performance Data

Student Theory Overview

One of the pages for the prototype (see Figure 23) presents an overview of the theory activity of the user. Last completed lessons visualizes an overview of the last lessons the student successfully completed. A filtering option for specific completed lessons is provided to give the user an option to view total time spent, the preferred time of day for interaction, and the average time spent on a page. A clock-graph visualization is used to show the most active theory interaction time. Descriptive feedback about the data may give the user information about new lessons, time since last interaction, provide lesson recommendations based on game performance, and give some feedback on performance after interacting with related theory.

Student Drill Overview

A dashboard prototype page for drill activity can be seen in Figure 24. Last completed drills provides an overview of the last drills that the user successfully completed. Total Drill Overview Stats gives the student feedback on the total number of drills completed, the completion rate (percentage of drills that have been successfully completed) and total drills scheduled for the course. A student may complete more drills than what is scheduled. As in the theory overview, the drill overview also has a clock-graph visualization to show the most active drill interaction times. Descriptive feedback can provide recommendations for drills (e.g., journal topics, game performance, "higher completion rate equals higher exam results").

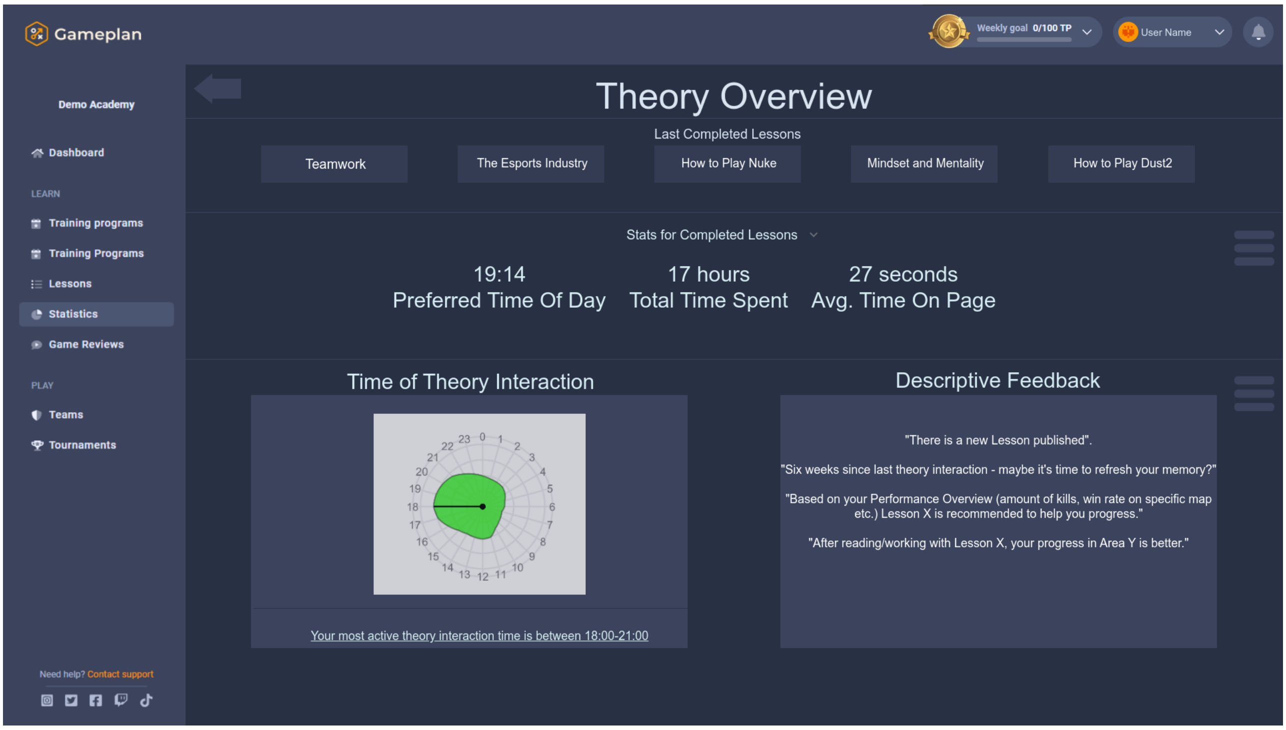

Figure 23: Prototype Image 3: Theory Data

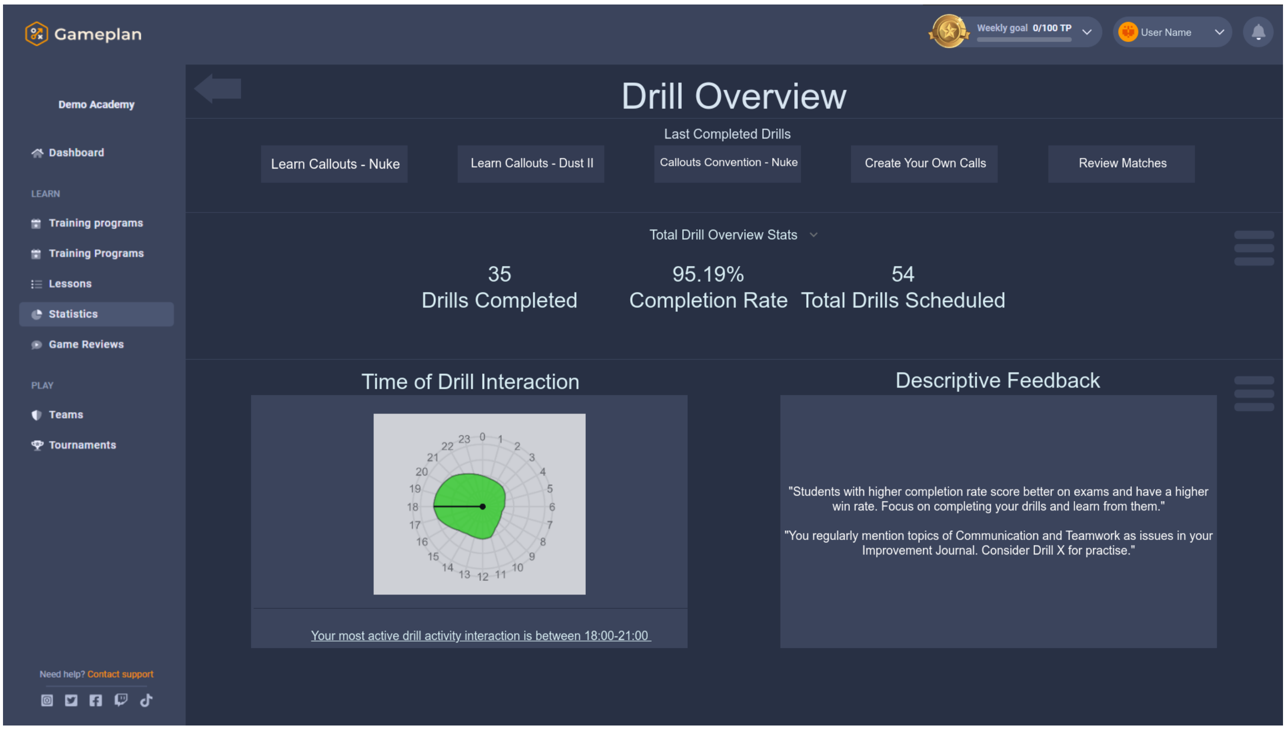

Figure 24: Prototype Image 4: Drill Data

Student Journal Overview

Figure 25 is a prototype page for a journal overview. The page produces an overview of the last journal entries, labelled by date of entry. A line graph visualization brings an overview of the mood trend for the specific user, displayed with date values along the x-axis, and mood values on the y-axis. A journal wordcloud visualizes words that are used in the journals, along with their frequency (bigger words equals a higher frequency). In combination with descriptive feedback, this page gives students insight into their journal use, their mood trends, and recommendations or descriptive insight into the data for assessment.

Figure 25: Prototype Image 5: Journal Data

Student Trend Overview

Combining datasets to visualize trends is one of the intended functionalities for the prototype (see Figure 26). Trends can be any combination of data to give valuable feedback and assessment for the user (e.g., recommended session length, journal impact, most effective time to play, most successful in game roles). Users may edit the trend overview to create awareness into the features they want.

Student Leaderboard Overview

The last page that is designed for the prototype is a leaderboard overview (see Figure 27). By providing a general outlook through leaderboards, the system can deliver a method of comparison to other students and placing their own performance in relation to a class average.

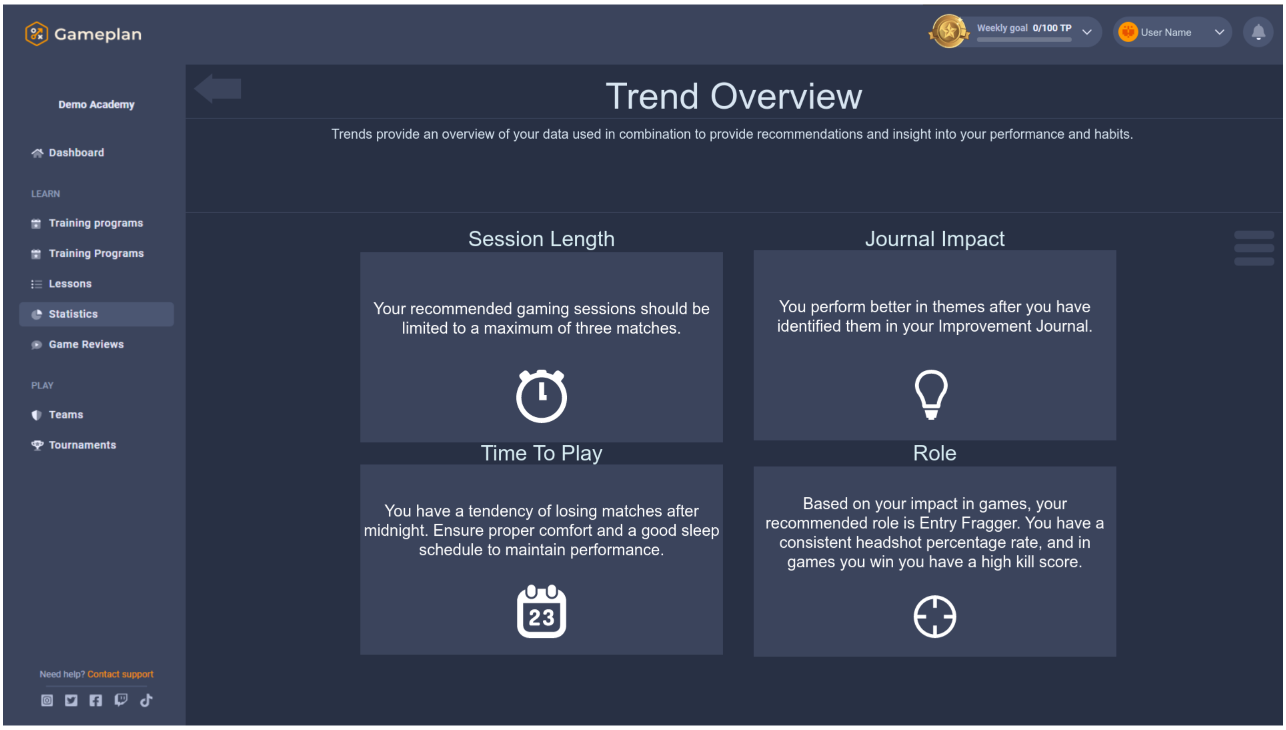

Figure 26: Prototype Image 6: Trend Data

Rank	Student	Win Rate	Theory Entries Read	Drills Completed
	Student Name	59.40%	211	23
2	Student Name	59.21%	245	25
3	Student Name	58.84%	169	11
4	Student Name	57.23%	174	7
5	Student Name	56.12%	64	26
6	Student Name	53.12%	0	13
7	Student Name	52.99%	133	5
8	Student Name	51.24%	214	14
9	Student Name	49.69%	77	17
10	Student Name	49.01%	145	2

Figure 27: Prototype Image 7: Leaderboard Data

6.2 Summary

The prototypes that were presented in this chapter can serve as inspiration to be used by Learn2Esport to improve their platform in terms of creating feedback to students on their activity and performance. The visualizations were based on theory from learning analytics dashboards and open learner models on how to present data in student-facing dashboards. Certain functionalities are results of data triangulation (combination of distinct datasets) that may reveal hidden patterns.

7 Discussion

This chapter discusses the analysis of data and prototype development, and presents the findings in relation to the research questions.

7.1 Research Approach

This research was guided by disciplines and principles of Learning Analytics (Long, 2011) and a multitude of methods were used as described in Chapter 3. The *desk research* phase was based on preliminary analysis and data processing to effectively identify limitations to the data, and to aid in the construction of research questions. A domain expert from Learn2Esport helped describe the data and its properties, and access to the platform was provided as part of *exploratory analysis*. This was an essential part of the research to identify different user groups based on the functionalities of the platform. In conjunction with the exploratory analysis, it was crucial to combine the findings to literature on learning theories and concepts. A literature review with focus on esports, learning theories, and student-facing dashboards was put together to implement this idea. There is a clear gap in the use of learning concepts when developing student-facing dashboards, but *self-regulated learning* is the most commonly used (Jivet et al., 2017). Learn2Esport (n.d.) state that they focus on *"equipping people with 21st-century skills through esports"*, hence the inclusion of 21st-century concepts in the research. As the focus of 21st century skills focus on much of the same concepts as self-regulated learning (e.g., *learning how to learn*), self-regulation concepts were also included as it is a more mature field with extensive research spanning decades. This combination of learning theories contributed to a strong foundation for the data analysis and prototype development. Thus, the research was guided by pedagogical purposes as opposed to only focusing on data availability (Booth, 2012; Kivimäki et al., 2019).

The data in this research originate from Learn2Esport, and all data properties are anonymous. Prior to the end of 2020, access to preliminary data was shared for use in preliminary analysis (originating from 2019). All available data was included in the finalized dataset, however, the data from 2019 was omitted from analysis due to risk of bias. Journal entries from 2019 had been analyzed and processed multiple times in the *thematic analysis* phase. Due to inclusion of three unique coders for the journals, the coding phase was considered to be more accurate if each coder worked with data that they had not seen beforehand. This created a natural span to limit the dataset to only include data from 2020.

Visualization of student data was based on literature from *learning analytics dashboards* and methods of *opening the learner model*. As recommended by Bodily, Kay, et al. (2018), academics should include both of these disciplines when working with student-facing dashboards as both methods focus on a common goal, but use different techniques to reach them. Inclusion of both disciplines laid the ground for an approach to the prototype visualizations that focused on both user modeling (which is common for OLMs) and a more data focused incentive (more common for LADs, and often with other goals than just focusing on the learner model).

The combination of learning concepts and data handling techniques yielded a unique research opportunity that: (1) included learning theories in the approach to data analysis and visualizations with focus on pedagogy, (2) produced an opportunity for esports students to gain insight into their data, and receive feedback on their performance and trends, and (3) provide prototype ideas for new functionalities for the development of esports management systems based on the data that is produced by students.

One of the novelties in the research was the inclusion and analysis of unstructured textual data from an educational esports environment. The journal functionality of Learn2Esport is used as a measure to reflect on performance in terms of how the game went. It is not uncommon to study self-report data in terms of surveys or structured entries to reflect on performance, but unstructured data requires processing to be further analyzed. Thematic analysis was applied to the preliminary data to organize the data and create a codebook, and this was based on lessons created by Learn2Esport and sentence-level themes from the preliminary journals. The combination of raw performance data from the games of the students in combination with themes extracted from journal entries can supplement performance data with the students' own interpretation of their performance.

7.2 Activity Data

Section 5.1.1 presents an overview of the profile linking percentage rate for all schools that offer esports curriculum through Learn2Esport. Considering that profile linking is a voluntary action, three out of four schools had a relatively high profile linking rate where the lowest of the three showed 46.4%. The last school in the group, labelled School D, had a 17.6% profile linking rate. As linking an account is voluntary, there might be lack of focus on providing adequate information about the platform and its functionalities. There is reason to believe that a designated introductory phase with lessons to provide guidance to the platform may result in a higher profile linking rate for the schools. Other issues such as teacher encouragement or lack of understanding may also be impactful, thus providing a safe and knowledgeable work environment for teachers is important.

Looking at theory activity statistics from Section 5.1.2, comparisons were made between the distinct groups to discover differences in theory frequency and statistics. The group that was defined as *students who use both concepts of self-monitoring and self-evaluation* N_{JE} had the highest number of average pages read per student. Subsequently, the group that was defined as *students who used the concept of self-monitoring* N_{PL} had the second highest number of average pages read per student.

The analysis revealed that the most active students in terms of theory activity were the ones who practiced both SRL-concepts. Section 5.2.1 provided visualizations of theory interaction frequency and revealed that students who apply SRL-concepts are more consistent and more active between the months, and also showed that the *smallest group of students had 89% of all theory interaction for August*, and that group is comprised of students who use both self-monitoring and self-evaluation.

In terms of theory interaction alone, it is difficult to state that *more is better*, however, a higher number of theory interactions equaled more lessons completed. By calculating central tendency it was possible to look into the affect of outliers that could skew the data. All groups revealed low values, disclosing that none of the groups were heavily affected by outliers in the data.

In Section 5.1.3 there were comparisons of drill activity. The group with the highest participation rate were the students who used self-monitoring and self-evaluation. Compared to theory activity, the number of students who did not complete any drills were much greater, hence the calculation of a participation rate. All user groups except for the group of students who did not use any of the defined SRL-concepts had a participation rate above 50%.

Looking at the average number of drills per student and the completion rate revealed that the students who only used self-monitoring scored higher than students who applied both concepts. Having a higher completion rate implies that the group completed more drills in terms of how many were started. Again, as with theory activity, a higher number of drills does not imply that more is better, but it reveals that the usage of drills were more common. Drills are not mandatory, but they are part of lessons as a method to reinforce learning, or put theories and new knowledge into practical action. Having first interacted with theory, then receiving instruction for how to perform certain actions can prove useful in performance development and practicing behaviors. Findings revealed that drill activity results were relatively close between all the user groups that are defined to use any variation of self-regulated learning concepts. This can also be seen in Section 5.2.2 in the visualizations that revealed drill frequency. All student groups see variations in terms of drill frequency between the months, but the only group that has decreasing drill participation for all months is N_{NPL}. The group with the lowest average number of drills, lowest participation rate, lowest completion rate, and being most affected by outliers is N_{NPL}, which is defined as students that did not apply any of the defined SRL-concepts.

Interpreting the exam results from Section 5.1.4 reveals that students who used both SRL-concepts had a total exam score percentage higher than all the other groups. This group also had the highest number of exams started per student. The completion rates between the groups only differed by 2.14% between the lowest and the highest. There were multiple issues identified with the exam activity data properties, as mentioned in Chapter 4. Despite this, the completion rates between the groups were very close, and did not reveal significant differences.

The exams are not final grades that students receive at the end of the course. Instead, they are tests that students complete after certain lessons, and are often scheduled in their calendar. Thus, the exams are not indicators as to course performance, but they do reveal a correlation between the usage of SRL-concepts, as the group that applied both concepts scored higher on the exams, read more theory, and completed the most drills.

The development and calculation of *the proxy variable for time spent on theory page* is described in Section 4.6, and results are presented in Section 5.1.5. As the dataset lacked information about time spent on theory pages, the proxy variables were established to reveal differences in time spent between the groups. All groups with students that applied any of the defined SRL-concepts had an average time spent per page between twenty and twenty-five seconds. Students that did not apply any of the SRL-concepts had an average time per page of thirty-five seconds. The findings revealed that students who used self-monitoring and self-evaluation had a lower time spent per page on average, as well as a considerably lower number of entries that exceeded five minutes.

Time on task is an individual factor, and spending more time does not directly indicate worse performance, or the other way around. However, with the huge growth in entries that exceed five minutes for group N_{NPL}, it is reasonable to indicate that these students may have been less focused or had more trouble working with the theory. Group N_{NPL} is also the only group with a coefficient of variation value that is considered high, revealing that the group is less consistent, and more affected by outliers.

7.3 Performance Data

Section 5.3 presents results of performance comparison. Performance data was limited to students who linked their accounts, or in other words, applied the self-monitoring concept. The results are thus comparisons between students who *also* applied self-evaluation against students who did not. For CS:GO, findings revealed that the group of students who used self-evaluation were mainly new students or new accounts as 79.47% of all matches for that group were played during enrollment. The students that did not write journals had an enrollment rate of 55.44%. It may be easier for new players to use intended functionalities from the esports management system as they are not as familiar with the game as more experienced students are. The results also showed that the win rate for the self-evaluating group was higher between the two groups, despite their lack of experience. Looking at findings for LoL, the students that used self-evaluation concepts had a lower enrollment rate at 39.71%, compared to students who did not, at 41.72%. The majority of LoL matches were played *prior* to enrollment for all students. Compared to the CS:GO findings, the self-evaluating students for LoL had a lower enrollment rate. Despite this, the win rate was still higher for the self-evaluating group. The findings indicate that the group of students who had the most theory interaction, highest exam scores, highest drill participation rates, and lowest average time spent on theory did better in terms of winning matches during esports enrollment. This result was present for both games.

Analysis of individual match performance (see Section 5.4) included the top five most active students from each game in terms of matches played. The exclusion of students in this analysis stage was decided by: (1) ensure many matches to neglect the effect of randomness or luck, (2) include students from both games, (3) as esports in education is a new discipline, there will be limitations in the curriculum, platform technology, teacher engagement, and other factors that may explain low activity from specific students. Worth mentioning is the current use of esports courses in the Swedish high schools: the course is not graded, and neither exams nor curriculum activities are mandatory. The analysis of the most active students demonstrates the value that rich esports data can provide, and should be treated as indicators to further develop curriculum and technology to continuously promote student encouragement.

7.4 Research Questions

RQ1: How are self-regulated concepts related to esports performance and activity?

There were a total of five user groups identified in the dataset, and they were based on two concepts from self-regulated learning; *self-monitoring and self evaluation* (Zumbrunn et al., 2011). This does not imply that students who were identified within a certain group are *determined* to be using the concepts (or not using them at all), but the platform provides an opportunity to use technology that can support such processes. Perry and Winne (2006) also claim that all learners are self-regulating, but not all are equally effective. The effectiveness can thus be distinguished by use of the platform functionalities.

One of the research inquiries was to observe any differences in activity and performance based on the user grouping, which then could be explained by the portrayal of self-regulation processes (or lack thereof) defining that student. It is evident that students who used self-regulated learning concepts had higher activity levels and performed better than students who did not.

All measures for activity data have shown that students defined as using self-monitoring and/or self-evaluation had higher activity levels, scored higher on exams, worked faster with the theory, and had higher completion rates than students who did not. In terms of match performance, students who used *both* SRL-concepts had the higher win rate for both games included in the study. The findings have revealed positive indications to focus on student participation in terms of profile linking (self-monitoring) and journal usage (self-evaluation).

RQ2: What can learning analytics tell us about esports students?

Through the use of learning analytics on esports data, we can make these observations about esports students:

- Concerning theory (or curricular) use, students who enact SRL-concepts of self-monitoring and self-evaluation interact the most with theory.
- There is a higher chance for students to participate and complete drills (i.e., small tasks, often based on previous theory interaction) if they apply either self-monitoring or self-evaluating methods.
- Exam results are not good indicators of other activity levels or performance results.
- Time on task (time on theory page) may indicate correlations to other activity areas, where a higher time on task negatively affects results. Students who did not enact SRL-concepts spent the most time on theory pages, and also revealed lower scores or lower interaction in other aspects.
- Students who write journal entries on their performance (self-evaluating) win more matches, despite being newer to the game.
- For esports data, generalization of recommended playtime and number of matches is not effective. Personalized feedback or recommendation is essential as there is too much variety between students.
- The most common theme that the esports students write about in their journals on self-evaluation are aspects of solo performance. Following, the most commonly mentioned themes are game mechanics, game phases, and emotional regulation.

Further, a component of the LA cycle (Figure 5) is based on how to feed the data back to students. By reviewing research on student-facing dashboards it became clear that both learning analytics dashboards and open learner models should be included in the design and development of dashboard applications for students as both disciplines are contributing to the development of dashboards. Data should promote awareness of performance and trends, be represented in a way that makes sense for the subject, provide reflection in terms of recommendations and feedback, and yield some impact in terms of what actions to choose next. Visualization of data can be presented in many forms, but building an editable dashboard gives students the opportunity to customize it to their needs and preferences.

Journal analyses displays the opportunity to analyze unstructured text input from the students to provide meaningful feedback that correlates with other aspects of the learning platform, such as lessons or game specific data. Adding self-report data includes the students own interpretation of their performance, which is then processed to provide meaningful feedback to the subject. The inclusion of a student's own data can make students more comfortable and trusting in a system where they see the effect of their actions. As stated by Vilalta et al., for students to act on their behaviors and apply SRL concepts they must become aware of both their knowledge (through self-report data) and behavior (activity and performance data) (Vilalta et al., 2009).

RQ3: How are session lengths tied to performance?

It is challenging to determine a generalized recommendation or limitation as to *how long you should play* or *what number of matches are recommended for top performance*. Students have contrasting techniques and approaches, and what might be recommended for one student can be the opposite for another one. Results indicate that for most students, the month with a higher win rate had a different trend in terms of the number of matches played, session lengths, and the variety in session distribution compared to the worst month. From this data, it is then possible to omit the most recurring trend from the worst month, and focus on the trends that were seen in months with a higher win rate. Interpreting the generalized data from sessions in CS:GO revealed slight differences between the best and worst months. The most notable difference being that four-game sessions should be limited, and students should play the majority of sessions in two-match sessions. To exemplify the challenge of acting on generalized recommendations in this instance: for the five analyzed CS:GO students, two students could be told to focus on two-match sessions, one to focus on one or two match sessions, and two to stick to three-match sessions. The same results can be seen for the LoL students. Presenting generalized recommendations such as *X is too many matches, Y is too few matches, Z is a balanced number of matches* will not always be applicable for every student, thus an individual analysis of the data to provide personalized recommendations and feedback is necessary.

7.5 Limitations

There were multiple limitations to consider for this research. The approach of applying concepts of self-regulation to some students does not, and should not, absolve other students of having the same traits. In terms of data availability and the functionalities presented in the platform, generalizations were made to differentiate students in terms of usage of these concepts. The population sample for the study is considered small, and the findings are mainly associated with individual activity and performance, thus generalization of the findings should be carefully considered.

Part of the study is likely susceptible to self-selection bias as the inclusion of students for the individual performance and session analysis was based on activity, and thus only included the top five most active students from both games in terms of matches played.

As both included video games are free to play, a student may create new accounts for the course, and be interpreted as a new player, thus creating false values for the enrollment rates in terms of experience. Limitations in the dataset are also worth mentioning considering there were empty entries, invalid properties, and functionalities that are in development on the platform. The development of a prototype design without evaluation of user subjects nor input from the esports management platform can be considered a limitation. Despite this, the prototype was designed for visualization of functionalities, and not meant to be any indication to what a final result should look like.

8 Conclusion

This chapter presents the conclusion and points to future work.

8.1 Summary

This research has demonstrated ways to handle student data that originate from the new and fast growing field of esports. By looking at esports in an educational setting, the focus has been to approach the data with pedagogical theories and student progression in mind. Applying learning concepts to the analysis of data can help differentiate students in terms of their activity levels and performance in matches. The use of self-regulating concepts can create better match performance and boost activity in terms of working on other curricular aspects, such as reading theory, completing drills, and score higher on tests. As esports is expanding and can provide both professional opportunities and relevant technical experiences, proper handling of student data and analysis of *how* and *why* progress occurs is important. It is essential to address advantages and challenges that follow such rapid development, and that is not always easy. Analysis should be based on pedagogical foundations, and not by data availability as is often the case with dashboard applications. Despite its rapid development it is important to realize that learning can be broken down to an individual level, thus the generalization of esports data is difficult in terms of recommending action and enhancing performance. Through proper design and analysis, esports data can prove invaluable in terms of progression and student learning.

The main contribution of this research is the use of learning analytics on student data that originate from an educational esports environment to provide insight into student behavior. Additionally, a literature review was conducted with focus on esports, learning theories, and student-facing dashboards. A mid-fidelity prototype was also designed based on the analysis results.

Based on the observed results of this research, the following conclusions are drawn: esports students who apply self-monitoring and self-evaluating concepts are more active when working with theory, completes more drills, scores higher on exams, and wins more matches in their respective games. These students are also more effective when working with theory in terms of time spent, and are less distracted when working through the curriculum. Personalized recommendations and feedback is needed to fully utilize the data that originate from esports students, although generalizations such as user grouping can provide valuable insight to other stakeholders and for curricular development.

Finally, the research shows the promise and use areas of esports data. There is a unique opportunity in such data rich environments to gain information about activity, performance, and concepts that may be limited in other areas. It is time consuming to process and analyze big amounts of data, but the creation of automated algorithms and machine learning methods could produce ad-hoc feedback that would provide valuable information for students and players in terms of game improvement and study progression.

8.2 Future Work

The next step in terms of the research findings would be to develop a functional prototype that can handle the student data and produce feedback based on the proposed design ideas where the students may edit their dashboards to their liking. The only missing step in the LA process cycle is the process of Feedback & Action, thus it is natural for future work to include that component. For both games included in the study, detailed information can be drawn from matches that have been played.

The nature of the analyses included in this thesis were purely descriptive and exploratory. Further analyses could involve inferential statistics to determine significant difference between the means of the groups, and correlational analyses.

It would be interesting to look at deeper data to identify patterns and variables and their impact factors on performance, and also including teammates and their contribution to the matches in terms of winning or losing. The implementation of an onboarding process could give students a proper introduction to the platform and also increase the number of students who link their accounts to the platform. This would create more student data and could create opportunities for new user grouping and deeper analysis.

Bibliography

Ahmad, N., & Bull, S. (2009). Learner trust in learner model externalisations. In *Aied* (pp. 617–619).

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning*, *20*(2), 13–29.

Bányai, F., Griffiths, M. D., Király, O., & Demetrovics, Z. (2019). The Psychology of Esports: A Systematic Literature Review. *Journal of gambling studies*, *35*(2), 351–365. doi: 10.1007/s10899-018-9763-1

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Office of Educational Technology, US Department of Education*.

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. *Journal of Computing in Higher Education*, *30*(3), 572–598.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41–50).

Bodily, R., & Verbert, K. (2017a). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418.

Bodily, R., & Verbert, K. (2017b). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 309–318).

Booth, M. (2012). Learning analytics: the new black. *Educause Review*, *47*(4), 52–53.

Braun, V., & Clarke, V. (2012). Thematic analysis. *American Psychological Association*.

Brem, A., Viardot, E., & Nylund, P. A. (2020). Implications of the coronavirus (covid-19) outbreak for innovation: Which technologies will improve our lives? *Technological forecasting and social change*, *163*, 120451.

Broadbent, J., Panadero, E., Lodge, J. M., & de Barba, P. (2020). Technologies to enhance self-regulated learning in online and computer-mediated learning environments. In *Handbook of research in educational communications and technology* (pp. 37–52). Springer.

Brun, A., Bonnin, G., Castagnos, S., Roussanaly, A., & Boyer, A. (2019). Learning analytics made in france: the metal project. *The International Journal of Information and Learning Technology*.

Bull, S. (2004). Supporting learning with open learner models. *Planning*, *29*(14), 1.

Bull, S., & Kay, J. (2007). Student models that invite the learner in: The smili:() open learner modelling framework. *International Journal of Artificial Intelligence in Education*, *17*(2), 89–120.

Bull, S., & Kay, J. (2010). Open learner models. In *Advances in intelligent tutoring systems* (pp. 301–322). Springer.

Bull, S., & Kay, J. (2016). Smili:() a framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, *26*(1), 293–331.

Chaubey, A., & Bhattacharya, B. (2015). Learning management system in higher education. *IJSTE-International Journal of Science Technology & Engineering*, *2*(3).

Colthorpe, K., Zimbardi, K., Ainscough, L., & Anderson, S. (2015). Know thy student! combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated learning. *Journal of Learning Analytics*, *2*(1), 134–155.

Davidian, M., & Louis, T. A. (2012). *Why statistics?* American Association for the Advancement of Science.

Denden, M., Tlili, A., Essalmi, F., & Jemni, M. (2018). Implicit modeling of learners' personalities in a game-based learning environment using their gaming behaviors. *Smart Learning Environments*, *5*(1), 29.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of interactive online learning*, *12*(1), 17–26.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 120–129).

Elias, T. (2011). Learning analytics. *Learning*, 1–22.

Ellis, R. A., Han, F., & Pardo, A. (2017). Improving learning analytics–combining observational and self-report data on student learning. *Journal of Educational Technology & Society*, *20*(3), 158–169.

Ellis, R. K. (2009). Field guide to learning management systems. *ASTD learning circuits*, 1–8.

ESPN. (2017). *Why the associated press stylebook went with esports, not esports.* https://www.espn.com/esports/story/_/id/19860473/why-associated-press-stylebook-went-esports-not-esports. (Accessed 05-2021.)

eSportsNS. (n.d.). *esports in nordic schools.* https://esportsns.com. (Accessed 05-2021.)

Freire, M., Serrano-Laguna, Á., Manero, B., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: learning analytics for serious games. In *Learning, design, and technology* (pp. 1–29). Springer Nature Switzerland AG.

Fromme, J. (2003). Computer games as a part of children's culture. *Game studies*, *3*(1), 49–62.

Gaftandzhieva, S., Docheva, M., & Doneva, R. (2020). A comprehensive approach to learning analytics in bulgarian school education. *Education and Information Technologies*, 1–19.

Gameplan. (n.d.). https://gameplan.com. (Accessed 05-2021.)

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113–128.

Grammarly. (n.d.). *What does "meta-" mean.* https://www.grammarly.com/blog/meta-meaning/. (Accessed 05-2021.)

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42–57.

Hamari, J., & Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research*, *27*(2), 211–232. doi: 10.1108/IntR-04-2016-0085

Hulaj, R., Nyström, M. B., Sörman, D. E., Backlund, C., Röhlcke, S., & Jonsson, B. (2020). A motivational model explaining performance in video games. *Frontiers in Psychology*, *11*.

IESF. (n.d.). *International esports federation.* https://www.ie-sf.org/. (Accessed 05-2021.)

ISO/IEC-JTC. (2013). *Sc36.* https://web.archive.org/web/20130522095557/http://www.sc36

.org:8080/. (Accessed: 05-2021.)

Jenny, S. E., Manning, R. D., Keiper, M. C., & Olrich, T. W. (2017). Virtual(ly) Athletes: Where eSports Fit Within the Definition of "Sport". *Quest*, *69*(1), 1–18.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *European conference on technology enhanced learning* (pp. 82–96).

Kivimäki, V., Pesonen, J., Romanoff, J., Remes, H., & Ihantola, P. (2019). Curricular concept maps as structured learning diaries: Collecting data on self-regulated learning and conceptual thinking for learning analytics applications. *Journal of learning analytics*, *6*(3), 106–121.

Learn2Esport. (n.d.). https://learn2esport.com. (Accessed 05-2021.)

Lee, J. S., Wu, M., Lee, D., Fleming, L., Ruben, L., Turner, T., … Steinkuehler, C. (2020). Designing an interest-based integrated curriculum around esports. *International Journal of Designs for Learning*, *11*(3), 78–95.

Long, P. (2011). *Lak'11: Proceedings of the 1st international conference on learning analytics and knowledge, february 27-march 1, 2011, banff, alberta, canada.* ACM.

Macarini, L. A., Lemos dos Santos, H., Cechinel, C., Ochoa, X., Rodés, V., Pérez Casas, A., … Díaz, P. (2020). Towards the implementation of a countrywide k-12 learning analytics initiative in uruguay. *Interactive Learning Environments*, *28*(2), 166–190.

Macon, B., Macon, L., & Phillips, N. (2016). Learning analytics: Case studies at a large urban community college. In *2016 international conference on computational science and computational intelligence (csci)* (pp. 258–263).

Manovich, L. (2010). What is visualization. *paj: The Journal of the Initiative for Digital.*

Misiejuk, K., & Wasson, B. (2017). State of the field report on learning analytics. *Centre for the Science of Learning & Technology (SLATE), University of Bergen.*

Mobafire. (n.d.). *League of legends fundamentals.* https://www.mobafire.com/league-of-legends/build/fundamentals-league-of-legends-new-player-beginner-guide-371292. (Accessed 05-2021.)

Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 33–44.

Nagyl, G. (2016). *Use of eye tracking for esports analytics in a moba game.* http://bora.uib.no/handle/1956/15875. (Department of Information Science and Media Studies, University of Bergen)

Newzoo. (2019). *Key global trends esports market report 2019 global free version.*

Newzoo. (2020). *Key global trends esports market report 2020 global free version.*

Pardo, A., Ellis, R. A., & Calvo, R. A. (2015). Combining observational and experiential data to inform the redesign of learning activities. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 305–309).

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A case study. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 422–429).

Pedraza-Ramirez, I., Musculus, L., Raab, M., & Laborde, S. (2020). Setting the scientific stage for esports psychology: A systematic review. *International Review of Sport and Exercise Psychology*, 1–34.

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gstudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review*, *18*(3), 211–228.

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In *Development of achievement motivation* (pp. 249–284). Elsevier.

Riot. (n.d.-a). *Learn the basics.* `https://na.leagueoflegends.com/en-us/how-to-play/`. (Accessed 05-2021.)

Riot. (n.d.-b). *Riot games.* `https://riotgames.com`. (Accessed 05-2021.)

Röhlcke, S., Bäcklund, C., Sörman, D. E., & Jonsson, B. (2018). Time on task matters most in video game expertise. *PloS one*, *13*(10), e0206555.

Sancho, M.-R., Cañabate, A., & Sabate, F. (2015). Contextualizing learning analytics for secondary schools at micro level. In *2015 international conference on interactive collaborative and blended learning (icbl)* (pp. 70–75).

Sapienza, A., Zeng, Y., Bessi, A., Lerman, K., & Ferrara, E. (2018). Individual performance in team-based online games. *Royal Society open science*, *5*(6), 180329.

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Journal of Educational Technology & Society*, *17*(4), 117–132.

Schubert, M., Drachen, A., & Mahlmann, T. (2016). Esports analytics through encounter detection. In *Proceedings of the mit sloan sports analytics conference* (Vol. 1, p. 2016).

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., ... Dillenbourg, P. (2016). Understanding learning at a glance: An overview of learning dashboard studies. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 532–533).

Steamcharts. (n.d.). *Steamcharts - an ongoing analysis of steam's concurrent players.* `https://steamcharts.com`. (Accessed 05-2021.)

Suero Montero, C., & Suhonen, J. (2014). Emotion analysis meets learning analytics: online learner profiling beyond numerical data. In *Proceedings of the 14th koli calling international conference on computing education research* (pp. 165–169).

Thurmond, V. A. (2001). The point of triangulation. *Journal of nursing scholarship*, *33*(3), 253–258.

Tiedemann, C. (2004). Sport (and culture of human motion) for historians,an approach to make the central term(s) more precise. *IX. international CESH-Congress*(2003), 1–8. Retrieved from `www.kulturwiss.info/`

Tukey, J. W., et al. (1977). *Exploratory data analysis* (Vol. 2). Reading, Mass.

Valve. (n.d.). *Valve corporation.* `https://valvesoftware.com`. (Accessed 05-2021.)

Vansteenkiste, M., & Deci, E. L. (2003). Competitively contingent rewards and intrinsic motivation: Can losers remain motivated? *Motivation and emotion*, *27*(4), 273–299.

Varanasi, M. R., Fischetti, J. C., & Smith, M. W. (2018). Analytics framework for k-12 school systems. In *Data leadership for k-12 schools in a time of accountability* (pp. 206–233). IGI Global.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500–1509.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*, *15*(3), 133–148.

Vilalta, R., Giraud-Carrier, C., & Brazdil, P. (2009). Meta-learning-concepts and techniques. In *Data mining and knowledge discovery handbook* (pp. 717–731). Springer.

Volk, H., Kellner, K., & Wohlhart, D. (2015). Learning analytics for english language teaching. *J. UCS*, *21*(1), 156–174.

Voogt, J., & Roblin, N. P. (2010). 21st century skills. *Discussienota. Zoetermeer: The Netherlands: Kennisnet*, *23*(03), 2000.

Wagner, M. G. (2006). On the scientific relevance of esports. In *International conference on internet computing* (pp. 437–442).

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational psychologist*, *30*(4), 173–187.

Winne, P. H. (2017). Learning analytics for self-regulated learning. *Handbook of learning analytics*, 241–249.

Wooden, A. (n.d.). *How big data is revolutionising the future of esports.* `https://www.intel.co.uk/content/www/uk/en/it-management/cloud-analytic-hub/big-data-powered-esports.html`. (Accessed: 05-2021.)

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, *25*(1), 82–91.

Zumbrunn, S., Tadlock, J., & Roberts, E. D. (2011). Encourage self regulated learning in the classroom. *Virginia Commonwealth University*.

APPENDIX

A Data Processing Queries

Following is an overview of most queries used for data processing and analysis. Some mathematical calculations were excluded. Queries were also applied in combination with other queries.

- R = Range
- C = Cell
- V = Value
- CRIT = Criterion

Query	Definition
=SUM(R)	Sum R.
=MAX(R)	Print highest value in R.
=MIN(R)	Print lowest value in R.
=STDEV(R)	Calculate standard devation of R.
=COUNT(R)	Count number of entries in R.
=UNIQUE(R)	Print unique values from R
=COUNTIF(R, C)	Count values in R if equal to C.
=COUNTA(R)	Count cells containing information in R.
=AVERAGE(R)	Calculate average of R.
=MEDIAN(R)	Calculate median of R.
=MODE.MULT(R)	Calculate multiple modes of R.
=TRIMMEAN(R,V)	Calculate trimmed mean for R on value V.
=TRANSPOSE(R)	Transposes R of cells
=TIMEVALUE(C)	Convert time to numerical.
=(C1/C2)*100	Calculate ratio between two values. E.g., win rate.
=COUNTIF(R,<V)	Count when R<V. Used to convert date values.
=SUMIF(R1, CRIT, R2)	Sum R2 corresponding value if R1 meets criterion.
=COUNTUNIQUE(R)	Count number of unique entries in R.
=COUNT(FILTER(R1,R2=C))	Count R1 by filtering R2 on value C.
=SPLIT(C," ", TRUE, TRUE)	Split C on character " ".
=FILTER(R1, MATCH(R2, R3, 0))	Query values on R1 that match between R2 and R3. Return index 0.
=CONTIFS(R1, V1, R2, V2)	Count if R1=V1 AND R2=V2.
=LEFT(C1,LEN(C1)-3),"")	Remove last 3 values from C1.
=IF(C2=C1,C4-C3,"")	Calculate time difference between two values.
=INDEX(R1,MATCH(C2,R2,0)	Return content of cell based on C2 match on R2, with offset 0.
=IFS(C1,CRIT1(R1),"Action1",C1,CRIT2(R1),"Action1"),"Action2")	Mark cell based on comparison.
=QUERY(R1, "select M, N, P, H, I, J, K, L, where M >= date '"	Part 1.
&text(date(R),"format")&"' and M < date '"&text(R2,"format")&"'")	Part 2: Query based on specific date values.
=IF(AND(R1=R2,R3-R4<V),R3-R4,"End")	Mark end of theory block if different date or time > 5 minutes.
=ARRAYFORMULA(TEXT(SUM(IFERROR(TIMEVALUE(R))),format))	Calculate time spent for R.