Teamwork assessment in cooperative video games:
Evaluating patterns, associations, and validity

by

Yvonne Farah

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:
Michael Dorneich, Major Professor
Jonathan Kelly
Stephen Gilbert

Iowa State University

Ames, Iowa

2023

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Michael Dorneich, and my committee members, Dr. Jonathan Kelly, and Dr. Stephen Gilbert, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my friends, colleagues, the department faculty, and staff for making my time at Iowa State University a wonderful experience. Moreover, I thank Games2Work’s team, who provided a supporting and engaging environment to conduct game-oriented research. I would also like to thank all the undergraduate students who worked with me throughout the project and provided significant help in designing and iterating the annotation process: Antonio Banuelos Moriel, Emma Edgar, Eli Whitney, and Thomas Walker.

ABSTRACT

This thesis describes the process of developing a teamwork measurement system by investigating cooperative video game environments for teamwork assessment. The work aims to provide teamwork measurement and testbed design guidance through cooperative games.

Teams are relied on in a variety of fields, due to the increased complexity of tasks, and the need for interpersonal cognitive, behavioral, and affective processes. However, quantifying teamwork behaviors is a challenging process, due to their complexity and interpersonal nature. Teamwork testbeds offer task environments for teamwork research, measurement, and training. Testbeds require the ability to elicit observable teamwork behaviors that can provide evidence of the teamwork competencies. This study investigates cooperative games as teamwork testbeds. Cooperative video games have been shown to foster prosocial behaviors, communication, and cooperative activities. Additionally, commercial off the shelf video games have been used as simulations for teamwork training and assessment. However, work is needed to establish the relationship between cooperative features and teamwork behaviors and investigate the internal validity of these environments as assessment testbeds. This thesis addresses these challenges through three research questions: investigating the consistency of cooperative games in inducing teamwork behaviors, the associations between cooperative features and teamwork behaviors, and the teamwork behavioral differences between upper and lower performers. With a clearer association between cooperative features and teamwork behaviors, designers can target specific profiles of teamwork behaviors in their testbed design. A codebook of teamwork behaviors and cooperative features was developed and used to annotate gameplay footage from publicly accessible streaming platforms. The study annotated footage from 177 teams, across 18 video games categorized under four cooperative genres.

Consistency of cooperative genres in inducing teamwork competencies was supported through similar distributions of teamwork behaviors within genres, for most competencies. In contrast, competencies between genres were significantly different. Therefore, designers can use cooperative genres to target teamwork distributions of interest. Additionally, associations between cooperative features and teamwork behaviors were established within genres, by investigating the top three competencies associated with every cooperative feature. The associations serve as testbed design guidance, where cooperative features can be used to emphasize the top associated competencies. Finally, when comparing upper and lower performers based on their gaming outcomes, some significant frequency differences were observed, for situation assessment, and explicit and implicit coordination, indicating that the frequency of some behavioral markers affect performance in several genres.

This study investigated cooperative gaming genres to explore how and why they induce teamwork behaviors. Cooperative genres were found to be inducing different patterns of teamwork competencies, and therefore can be used by designers in teamwork testbed developments, to target specific teamwork profiles. Furthermore, cooperative features induced different proportions of teamwork behaviors, and therefore can be provided as tools to make certain behaviors observable and trackable. The findings support the validity of these environments for teamwork assessment and provide a range of assessment and design insights to develop and assess teamwork in cooperative testbeds, including the developed codebook, the consistent teamwork profiles, and the derived associations.

CHAPTER 1. INTRODUCTION

Objective and Problem Statement

Objective

This thesis describes the process of developing a teamwork measurement system by investigating cooperative video game environments for teamwork assessment. The study investigates how and why cooperative games encourage teamwork behaviors, through exploring how consistently they induce similar distributions of teamwork competencies. Additionally, the work aims to establish associations between cooperative features and teamwork behaviors, to provide guidance for cooperative testbeds development. Through this investigation, the thesis intends to propose teamwork measurement and testbed development guidance, through cooperative video games.

Problem Statement: Teamwork Measurement

Teamwork assessment is a challenging process due to difficulties in quantifying complex interpersonal processes (Boyle et al., 2011), that can be cognitive, behavioral, and affective (Salas et al., 2005). As task environments become more complex and more dynamic, organizations rely on teams, in a wide variety of fields (Roberts et al., 2022), including emergency medicine (Rosen et al., 2008), military and aviation (Goodwin et al., 2018), sports (Carron et al., 2002), and collaborative learning (Zea et al., 2009). Therefore, inefficient teamwork can have negative economic, social, and organizational consequences (Roberts et al., 2022), which raises the need for teamwork measurement systems to achieve an enhanced understanding of team attributes, and how to foster effective teamwork processes (Roberts et al., 2022; Salas et al., 2005, 2008; Wiese et al., 2015). However, to reach a finer understanding of how teamwork skills develop, there is a need for measurement systems that can quantify behavioral changes (Kendall & Salas, 2004). Therefore, teamwork measures that can track and capture team dynamics as they evolve within the environment, provide an understanding on how teamwork skills develop and how they are affected by the collaborative technological environments (Marlow et al., 2016).

Problem Statement: Testbed Requirements

Designing task environments for teamwork assessment requires replicating the complexity and dynamic nature of situations that require people to work together as a team (Cooke et al., 2020). This has been explored through developing simulations that replicate real life situations (Marlow et al., 2016). Synthetic collaborative environments can effectively induce teamwork behaviors if they successfully recreate the targeted task cues and consequences (Bowers & Jentsch, 2001). This study aims to use cooperative video games to explore gamified teamwork testbed requirements, to systematically design environments that can elicit observable teamwork behaviors. Cooperative video games have been studied as a social interactions testbed and their mechanics have been proposed as capable of inducing collaborative activities (Coovert et al., 2017). However, to use them as teamwork testbeds, work is needed to establish associations between their design elements and teamwork behaviors (Marlow et al., 2016)

Problem Statement: Cooperative Games for Teamwork Assessment

Cooperative video games have been observed to foster social interactions, ranging from prosocial behaviors (Depping & Mandryk, 2017; Harris, 2019) to communication, leadership, and planning (Chen, 2009; Grandi, 2021; Jang & Ryu, 2011; Lisk et al., 2012; Peña & Hancock, 2006; Pobiedina et al., 2013; Williams & Kirschner, 2012). Additionally, they have been gaining attention in teamwork training and measurement (Marlow et al., 2016; Mayer, 2018 ; Coovert et al., 2017), where team science researchers have started investigating video games as suitable testbeds for teamwork training and assessment (Alexander et al., 2005; Belanich et al., 2004;

Fletcher et al., 2006; Mayer, 2018; Zhang et al., 2011). Which provides the opportunity to explore these environments to understand their consistency, and the relationship between the gaming attributes and teamwork behaviors (Marlow et al., 2016). Additionally, as the use of collaborative technologies for teamwork activities evolve, so is the need to understand what factors drive individuals to cooperate in these environments, and what design features can support and cultivate teaming behaviors (Morschheuser et al., 2017).

Potential Benefits

The study aims to address these challenges by proposing a teamwork measurement system that relies on tracking behaviors in cooperative gameplay footage and associating these behaviors with the cooperative features inducing them. Through this approach, three benefits are proposed. First, by quantifying behaviors, we aim to unveil patterns of teamwork competencies in cooperative games. By understanding the distributions of teamwork behaviors in games, and whether there exists a repetitive pattern that is relevant to the gaming environment, the study provides empirical teamwork profiles, that provide a better understanding of how teamwork competencies are influenced by the cooperative environment, and how to use cooperative games as testbeds to induce the targeted teamwork profile. Furthermore, the study invesitagtes the consistency of multiple video games and multiple teams within every genre, to provide findings that are representative of every genre, and therefore more generalizable in terms of testbed design guidance. Second by developing associations between cooperative features and teamwork behaviors, the study provides empirically based design associations, that can be used to trigger desired teamwork behaviors, and measure them accordingly, through cooperative features. And finally, the study investigates validity aspects of using video games for teamwork assessment, an emerging field in industry and literature. By exploring video games' validity, the study aims to answer whether cooperative features affect teaming behaviors, whether they are consistently induced and how can they be influenced through design decisions.

Research Questions

To explore the proposed benefits and address the described challenges, this work investigates three research questions (RQs).

RQ1: Are Cooperative Game Genres Consistent in Inducing Teamwork Behaviors?

This study separates cooperative games into genres, based on the dominant cooperative design features that are driving players' interactions with the game environment and other players (Wendel & Konert, 2016). This question aims to explore whether video games with similar cooperative features induce homogenous distributions of teamwork competencies. Through exploring this question, the researchers aim to understand whether genres within gamified environments can be used to systematically induce teaming behaviors. With distinct teamwork profiles within a genre, testbed designers can choose the genre that produces the distribution of teamwork behaviors they want in their testbed.

RQ2: What are the Associations Between Cooperative Features and Teamwork Behaviors?

In addition to studying whether consistent distributions of teamwork competencies can be induced by video games within the same cooperative genre, the study aims to understand what is driving the teamwork behaviors in the analyzed cooperative games, by annotating behaviors in association with the cooperative features inducing them, to develop an enhanced understanding of the relationship between game attributes and teaming competencies, and therefore provide design guidance on how to use cooperative features to induce desired teamwork processes. Additionally, through studying associations in 18 video games within and across four different cooperative gaming genres, the study aims to provide representative associations between features and behaviors that can be generalizable to teamwork testbed design.

RQ3: What are the Differences in Teamwork Behaviors Between Upper and Lower Performers?

Existing teamwork models propose that teamwork processes are mediators to translate team inputs into team performance outcomes (Ilgen et al., 2005). To investigate this area, the third research question aims to answer whether there are differences in the frequency of teamwork behaviors between upper and lower performers, based on teams gaming performance, such as time to finish levels, errors, and scores. Through this question, the thesis aims to explore whether there are behavioral differences within every genre, between upper and lower performers, providing clarity on what teaming behaviors are contributing to the higher outcomes, within cooperative genres. Additionally, this research question contributes to the internal validity of video games as testbeds for teamwork assessment (Mayer, 2018). By investigating whether teamwork behaviors are necessary for teams to achieve higher performance outcomes, the study aims to provide support to the claim that teamwork behaviors are necessary in these cooperative environments, and therefore support this aspect of internal validity (Mayer, 2018).

Cooperative Games Validity as Teamwork Testbeds

In systems evaluation, validity is defined as "a confirmation through objective evidence that the requirements for a specific intended use or application of a system have been fulfilled" (Wilson et al., 2016, p. 03). Therefore, the use of cooperative video games for teamwork assessment, relies on their ability to create situations that elicit measurable actions, which provide information about the construct of interest (Mislevy et al., 2012; Levy et al., 2016). Hence, the study aims to explore the capability of cooperative video games in eliciting observable teamwork behaviors that serve as indicators of targeted teamwork construct. Designing games for assessment relies on tuning situations and interactions to generate evidence of the construct of interest (Mislevy et al., 2012; Levy et al., 2016).

While this work does not claim to explore all areas of the internal validity of video games for teamwork assessment, it attempts to focus on internal validity through different investigations. Construct validity aims to answer whether manipulating the testbeds' variables can generate detectable effects on the teamwork constructs (Jentsch & Bowers, 1998). This form of validity is explored through RQ1, which investigates if by changing cooperative genres, the distribution of teamwork competencies changes. Another aspect of validity is convergent validity (Jentsch & Bowers, 1998). This can be assessed by determining if the designer can elicit behaviors that are related to the construct of interest through design manipulations (Jentsch & Bowers, 1998). By exploring associations between cooperative features and behavioral markers through RQ2, the study aims to provide testbed design insights that can elicit behaviors related to the targeted teamwork construct. Finally, another aspect of convergent validity can be assessed by determining if the better performing teams show differences in teaming behaviors compared to lower performing teams (Mayer, 2018). By understanding whether teamwork behaviors in games positively influence team performance outcomes (RQ3), the work contributes to the internal validity of games as teamwork assessment environments.

In conclusion, through studying 18 video games, within and between four cooperative genres, the study aims to systematically study teamwork interactions and associations with cooperative features and attempts to provide representative findings, that can be further generalized and applied in teamwork testbed development.

Approach

To answer the research questions, the researchers followed an iterative approach to develop a codebook of behavioral markers and cooperative features, rooted in teaming literature, and synthesized with gaming literature. The codebook was used to analyze gameplay footage of commercial cooperative video games through coding publicly available video streams. The coder associated observable cooperative features with the identified behavioral markers in time stamped entries. The approach of using existing footage from commercial cooperative games was followed, due to its ability to allow researchers to observe naturalistic team interactions within genuine, elaborate, and high-fidelity gaming environments (Harris, 2019; Isbister, 2010). Furthermore, it has been suggested that ecological validity of social play research can be supported through employing fully designed games, studying individuals with existing social and gaming familiarity, and studying play interactions in contexts that are naturalistic and familiar for the players, instead of the laboratory (Harris, 2019; Isbister, 2010). Therefore, this study's approach benefits from these advantages, through studying player interactions in existing fully developed games, in naturalistic contexts.

The resulting annotations were analyzed to answer the research questions, by investigating consistency of teamwork competencies distributions within genres and exploring differences between genres. Additionally, associations between features and behaviors were analyzed to establish an understanding of the relationship between cooperative features and teamwork behaviors and expand their systematic use in cooperative gaming testbeds for teamwork measurement. Furthermore, comparisons between upper and lower performers were conducted to assess whether there are differences in the frequency of teaming behaviors between the two outcome categories.

Through these three research questions, the study addresses the internal validity of using cooperative video games for teamwork assessment and aims to propose the development of teamwork measures and testbeds through cooperative video games.

Thesis Outline

The thesis is organized as follows. Chapter 2 covers existing frameworks of teaming competencies and behaviors. Additionally, it overviews the theory of developing teamwork measurement systems, and empirical studies where teamwork was empirically measured. Next, the chapter covers cooperative gaming literature, through presenting work on cooperative design features and studies in commercial and academic cooperative video games. Covering cooperative gaming literature aims to provide ground for their suitability for teamwork assessment, through existing work tackling cooperative patterns, features, and players' interactions in multiplayer video games. Chapter 3 discusses the methodology process, detailing the codebook development, video game selection, obtaining inter-rater agreement and applying the codebook to annotate gameplay footage. Additionally, it details the analysis plan aiming to address the three research questions of consistency, associations, and performance outcomes.

The next three chapters address every research question separately by presenting the results and discussion associated with the research question. Chapter 4 presents the results investigating the consistency of teamwork competencies (RQ1) within genres and differences between genres, followed by a discussion elaborating on the findings. Chapter 5 presents the results and discussion on associations between cooperative features and teaming behaviors (RQ2). Chapter 6 presents the teamwork behaviors' comparisons between upper and lower performers (RQ3). Chapter 7 elaborates on the synthesis between the three research questions through exploring the validity of cooperative games for teamwork assessment. Chapter 8 provides detailed design insights, summarized through visual representations of the important design features of every genre. Finally, Chapter 8 concludes the thesis by summarizing the findings of the study, contributions, limitations, and future work.

CHAPTER 2. RELATED WORK

This chapter covers related work that informed the development and application of the teaming measurement system in cooperative gaming environments. This work draws on previous research in teamwork science where teamwork frameworks, measurement systems and training methods have been established, based on theoretical and empirical research. The chapter transitions from teamwork science to cooperative video games' research, to establish the commonality between these two fields. By covering existing cooperative gaming studies that investigate cooperative design features, patterns and players interactions, the literature review provides grounds on how cooperative games are suitable for teamwork assessment. The review establishes the current state of teamwork measurement and testbed development for teamwork assessment through cooperative video games.

Teamwork Assessment

Teams perform in dynamic and complex environments. As the complexity of tasks exceed the individual's capacity, organizations depend on teams to work on interdependent tasks toward a common goal (Salas et al., 2005). Teams perform in a variety of fields, including aviation, healthcare, military, and engineering (Salas et al., 2008). As the demands for teams increase, scientifically rooted guidance is needed, to understand the effectiveness of teamwork. Previous work provides several frameworks that aim to establish a teaming taxonomy and classify teamwork processes. Moreover, measurement systems are provided to capture teaming inputs, processes, and outputs, through subjective and objective measures. Finally, researchers and practitioners have implemented training techniques to develop teaming environments and employ measurement systems.

Teamwork Frameworks

Teams have been defined as two or more individuals, engaging in interdependent tasks to accomplish a common mission (Dyer, 1984; Salas et al., 1992). The Input-Process-Output (IPO) model was suggested as a classic team systems model, where inputs are transformed into outcomes through processes (McGrath, 1984). Inputs are the initial conditions affecting the team prior to its performance episode. They constitute of knowledge, skills, and attitudes of team members, in addition to environment and team characteristics (Ilgen et al., 2005). Outputs are the resulting outcomes of the team's performance episode that are relevant to the team task. Processes are the behavioral, cognitive, and affective interactions that transform teams' inputs into outputs (Ilgen et al., 2005; Marks et al., 2001). Among the behavioral processes, previous work distinguishes between task work and teamwork behaviors (Morgan et al., 1993; Crawford & Lepine, 2013). Task work constitutes of individuals' behaviors with technical characteristics, which contribute to task progress (Bowers et al., 1997; Pinelle et al., 2003). Teamwork behaviors are the interrelated actions and verbal exchanges that team members engage in collectively, to facilitate the achievement of their collective tasks (Salas et al., 2005). Therefore, to understand the effectiveness of teams, researchers have investigated the teamwork aspect of team interactions, by proposing a variety of frameworks and taxonomies.

Teamwork behaviors have been synthesized into two categories: regulating team performance and managing team maintenance (Rousseau et al., 2006; Wittenbaum et al., 2004). The following paragraphs explore the framework of teamwork behaviors synthesized by Rousseau et al., (2006). Under regulation of team performance, there exists four proposed functions of preparation, execution, evaluation, and adjustment (Frese & Zapf, 1994). To prepare for their work accomplishment, teams focus on activities of analysis and planning. Mission analysis is defined as the collective actions of interpreting and evaluating team's goals, through identifying tasks, environment conditions, and available resources (Marks et al., 2001). This is followed by planning, which is the activity of specifying an explicit course of action, that can include team members' responsibilities, the sequence and timing of executed sub-tasks, and the methods used for executing actions (Marks et al., 2001; Rosen et al., 2011; Rousseau et al., 2006).

After the preparation function, teams move to task-related collaborative behaviors for execution (Rousseau et al., 2006). These behaviors were categorized under three dimensions: coordination, cooperation, and information exchange. Coordination involves the integration of individuals' activities, to ensure that their tasks are sequenced and synchronized (Cannon-Bowers et al., 1995). Cooperation involves team members behaviorally collaborating to execute a task (Yeatts & Hyten, 1998). Finally, information exchange involves sharing information (Janz et al., 1997) and closed-loop communication (McIntyre & Salas, 1995). Team members transfer information through reporting updates on new task-related information, available resources, task demands and delays.

While in execution phase, teams engage in evaluation. To assess their progress, teams monitor their performance and their environment (Salas et al., 2005). Mutual performance monitoring is the activity of tracking team members' progress, by keeping track of team-mates' work, through observation and self-regulation (Marks & Panzer, 2004; Marks et al., 2001). This action enables teams to detect deviations and inadequacies and adjust accordingly (Marks & Panzer, 2004). Systems monitoring involves internal and external tracking. Internal systems monitoring involves keeping track of team generated resources (e.g., equipment, personnel, team-generated information). External monitoring involves tracking environmental conditions (e.g., new demands, organizational changes, risks, and environmental resources) (Marks et al., 2001).

In the final function under regulation of team performance, teams engage in adjustment behaviors. Through evaluation, teams are capable of exercising adaptability. Adaptability can involve team members re-iterating some of the previous functions, such as adjusting their plans, re-visiting their goals, or increasing the frequency of coordination (Burke et al., 2006; Rosen et al., 2011a; Rousseau et al., 2006). Moreover, they can also engage in adjustment behaviors that include backup behavior, intrateam coaching, collaborative problem solving and innovation. Backup behavior involves team members helping others in performing their roles (Rousseau et al., 2006; Salas et al., 2005). Performance monitoring is closely related to backup behavior, since team members would be capable of noticing overloads and aid accordingly (Salas et al., 2005). Backup behavior can include corrective guidance, behavioral assistance (e.g., performing the team members' task), and sharing resources and supplies (Marks et al., 2001; Porter et al., 2003). Intrateam coaching is another behavior under adjustment. It involves exchanging constructive feedback and learning from other team members to improve performance. This can include providing advice, suggestions, and instructions (Druskat & Kayes, 1998; Rasker et al., 2000). In addition, teams engage in collaborative problem solving when encountering unexpected demands or failures. This process is like mission analysis however it focuses on the encountered problems by gathering information, integrating, and identifying solutions (Rousseau et al., 2006). Finally, under adjustment teams can engage in innovative behaviors. These involve generating new approaches. In dynamic environments, teams might be required to continuously improve their methods and come up with new ideas and proposals (Kozlowski et al., 1999).

Teams also engage in behaviors to manage the team's maintenance (Rousseau et al., 2006). This dimension focuses on psychological support and conflict management. Psychological support encourages team members to express their concerns and deal with emerging stresses and pressures that arise when teams perform together (Rosenfeld & Richman, 1997). It can include motivation, raising team spirit, and supportiveness. Second, conflict management targets solving team conflict through a collective effort to understand team members' interests and resolve agreements (Rousseau et al., 2006). This behavior includes taking different team members' perspectives into consideration, and generating resolving decisions that aim to advance the teams' common goal (Rousseau et al., 2006).

Three coordinating mechanisms can facilitate team processes: mutual trust, closed loop communication and shared mental models (Salas et al., 2005). A distinction between emergent states (e.g., mutual trust, shared mental models) and processes (e.g., backup behavior, monitoring) is essential at this stage. Emergent states involve cognitive and affective team states rather than the explicit teams' interactions (Marks et al., 2001). Therefore, they are not team interactions nor actions that explicitly lead to team outcomes. Emergent states such as mutual trust and mental models are considered facilitating mechanisms (Salas et al., 2005). Shared mental models are the common structured perception of the teams' mission, task environment, and performance expectations (Cannon-Bowers et al., 1993). Therefore, shared mental models are developed through team processes described previously (e.g., systems monitoring, mission analysis, planning) and in-turn support these processes by serving as a shared mental framework that allows teams to execute their actions (Zaccaro et al., 2001). Therefore, shared mental models are formed and updated through the team's performance cycle. Mutual trust is the common perception that team members will engage in actions that are beneficial to the team (Webber, 2002). Therefore, while engaging in interdependent tasks, team members engage in a level of risk, where they rely on each other to execute the agreed-on plans, perform adequately, and cooperate toward the common goal (Salas et al., 2005).

Team adaptation has been proposed as a global property of the team's performance (Kozlowski et al., 1999). Following the IPO model, individual characteristics include Knowledge (e.g., Task expertise, Team expertise), Attitudes (e.g., Orientation toward team members) and Traits and Abilities (e.g., Cognitive capacities, openness to team inputs). These inputs support the processes, categorized under an adaptive cycle proposed by (Burke et al., 2006). These processes include situation assessment, where team members recognize cues relevant to the team's mission and communicate their meaning. Followed by plan formulation where teams engage in mission analysis and planning. Afterwards teams engage in execution including monitoring, communication, backup behavior and leadership. And finally, team learning where team members engage in reflective processes to review their performance and integrate new strategies in next cycles.

In conclusion, this chapter reviewed a variety of teaming frameworks, and explored a synthesized framework of teamwork behaviors proposed by Rousseau et al., (2006). Categorizing teaming behaviors into different functions was led by theoretical work and supported by empirical findings. There is a synthesis in literature that separates teamwork from taskwork (Salas et al., 2008; Wiese et al., 2015). Under teamwork, teams engage in a variety of behaviors to prepare, execute, and monitor their progress. Additionally, they go through adaptive cycles to develop and maintain mental models and shared situation awareness. Processes are separated from emergent states, as the latter is a cognitive formation that supports the team's behaviors, including mutual trust and mental models for example. Previous work suggests that empirical research is further needed to verify the conceptual structures proposed in these frameworks (Roberts et al., 2022; Wiese et al., 2015; Rousseau et al., 2006; Kendall & Salas, 2004). Moreover, valid, and reliable measures are needed, that are based on theoretical frameworks, but also provide insights on how these structures are applied in teams' performance episode. Hence, the next section of this chapter goes over teamwork measurement systems that aim to measure the proposed processes and emergent states described in the reviewed frameworks.

Teamwork Measurement Systems

A team measurement system includes a variety of approaches employed for teamwork observation and assessment (Wiese et al., 2015). A complete measurement system captures individual characteristics (Inputs), team processes and emergent states (Processes) and team outcomes (Outputs). Dimensions of a measurement system can be divided into two levels: team level and individual level (Wiese et al., 2015). Team level processes include team-specific knowledge, skills, attitudes and other (KSAO), teamwork processes and emergent states, and team performance outcomes. Individual level processes include Individual KSAO, taskwork processes and individual performance outcomes. Their work also detailed six considerations that support the design, development, and implementation of team measurement systems: purpose, constructs, referent, source, timing, and validity.

First, a measurement purpose needs to be specified to select relevant teaming constructs. A teaming construct is a measurable team process or emergent state (Wiese et al., 2015). Second, practitioners and researchers select the constructs of interest across the three stages of measurement (Inputs, Processes and Outputs). Third, the system needs a measurement referent. The measurement developer needs to consider the focal point of the measure, specifying whether the measure is individual oriented (e.g., tracking individual behaviors) or team oriented (e.g., tracking collective behaviors). Fourth, the system needs a measurement source. A measurement source can be the team members themselves (can access own cognition but can be subjective, too lenient, or too harsh), trained raters (trained to observe and assess behaviors) and automated sources (Wiese et al., 2015). Measurement sources have an impact on the validity of the measure. Specifically, subjective measures including surveys and interviews, and have cognitive access benefits (e.g., directly asking individuals about their teaming experience) but can be biased and inaccurate (Roberts et al., 2022; Sottilare et al., 2018). Fifth, the measurement system can target certain timings. As described in the teamwork framework sections, teams follow temporal cycles. Therefore, certain constructs are more suitable to measure at certain points of the cycle (Wiese et al., 2015). Finally, practitioners and researchers need to consider measurement validity to assess whether the measure is capturing what it was designed to measure (Roberts et al., 2022; Wiese et al., 2015).

Teamwork measurement tools

Measurement tools are used in a measurement system to assess team inputs, mediators, and outcomes. Measurement tools can be event-based, automated performance monitoring, behaviorally anchored rating scales (BARS), behavioral observation scales (BOS) and self-report measures (Kendall & Salas, 2004). Observational measures that do not rely on self-report and automated psychometric measures, rely on behavioral markers, which are trackable indicators of a teaming competency, such as observable and audible behaviors (Sottilare et al., 2018). BARS are tools used to classify and rates teaming behaviors, while BOS assesses the frequency of behaviors. The event-based measurement aims to create simulated teaming exercises, where designers embed trigger events to elicit targeted teaming behaviors (Fowlkes et al., 1998). Examples include the Event Based Approach to Training (EBAT) and the Targeted Acceptable Responses to Generated Events or Tasks (TARGETs) (Fowlkes et al., 1998; Fowlkes et al., 1994). Automated Performance Monitoring is a method that continuously monitors team members using computers, such as body movements and verbal communications. Finally, self-report measures capture constructs from individual team members' perspectives, using questionnaires for example. Self-report measures can provide access to emergent states and cognitive teaming constructs, such as trust, cohesion, and mental models.

In a recent review of teamwork measurement systems, (Roberts et al., 2022) highlight the frequent use of self-report and observational measures in literature. While self-report measures are valuable in accessing affective emergent states, such as trust, which are not directly observable, they are susceptible to inflated and biased scores (Marlow, et al., 2018; Sottilare et al., 2018). Additionally, since they are delivered at a static point in time (pre- or post- team task), they are incapable of capturing dynamic team processes, that evolve while the team is performing their tasks (Rosen et al., 2012). Alternatively, observational measures can capture the dynamic nature of teamwork behaviors, since they are employed in real-time or through recordings of the team activities (Roberts et al., 2022). Particularly, when employed accurately, capturing observable behaviors can provide direct assessment of team attributes (Salas et al., 2017). Additionally, observational methods provide a higher level of unobtrusive assessment compared to self-report measures and can reflect a more objective approach (Roberts et al., 2022). However, these methods are also subject to drawbacks, including bias and subjectivity, which should be mitigated through inter-coder agreement, rater coaching, and detailed scoring guidelines (Weaver et al., 2010).

Observational teamwork studies have been employed in a variety of fields. Evaluating teamwork faces difficulties due to its need to quantify inherently complex behaviors (Boyle et al., 2011), which can be mitigated through developing accurate behavioral markers to represent the fundamentals of teamwork attributes (Salas et al., 2017). Observational studies have been used to measure team cognition, through observing team processes (Cooke et al., 2013), and team situational awareness (Gorman et al., 2017). Behavioral markers'-based teamwork measurements have been commonly employed to study teamwork interactions in surgical contexts (Whittaker et al., 2015). Examples include the Observational Teamwork Assessment for Surgery (OTAS), where behaviors are analyzed in five areas: communication, cooperation, coordination, shared leadership, and team monitoring (Healey et al., 2004). Another example is the Nontechnical skills for surgeons (NOTTS), which relies on a behavioral marker system to measure four competencies of situation awareness, decision making, communication and teamwork (Yule et al., 2006).

Finally, automated measures are gaining attention in recent literature, due to their ability to capture a variety of teamwork processes (cognitive, affective, and behavioral), in addition to being unobtrusive and dynamic (Roberts et al., 2022; Gorman et al., 2012; Salas et al., 2017). Automated measures can capture physical properties of the behavior, such as frequency and duration, the content itself, and the sequential flow of behaviors (Kiekel et al., 2002). Therefore, providing a holistic approach to team measurement. However, as of now, human processing remains required, to derive meaningful teamwork metrics, and semi-automated approaches might be more appropriate (Granåsen 2018; Roberts et al., 2022). Nonetheless, automated technology requires rigorous construct validation, to establish connections between what is being measured and the targeted construct of interest (Braun & Kujanin 2015).

In conclusion, teamwork measurement systems and applied techniques vary depending on several factors described above. Assessing teams in dynamic and adaptive systems, require measures that are unobtrusive, real-time, and practically implemented (Salas et al., 2008; Roberts et al., 2022) . This highlights a limitation of self-report measures, since they require direct input from team members, and therefore cannot be implemented in real-time during the performance episode, hence the increasing interest in measures that can quantify teaming behaviors (Sottilare et al., 2018) . Furthermore, in recent literature more attention is being emphasized on developing unobtrusive methods that can directly assess the dynamic processes of teams (Roberts et al., 2022; Salas et al., 2017). Observational team measurement is one prominent field in literature, where observable behavioral markers are essential to model teamwork attributes (Salas et al., 2017).

The first two sections of the chapter covered theoretical and conceptual work in team science literature, reviewing existing frameworks and measurement techniques. The next section in this chapter transitions to cooperative video games, introducing them as suitable environments to apply the teamwork models and measurement systems described in this section, due to their shared goal structures, design features and patterns, and communication support.

Cooperative Video Games: Design and Research

Developing teamwork measurement systems requires testbed environments. Cooperative video games have gained attention as testbeds for social interactions studies, such as investigating how interdependence and shared goals foster prosocial behaviors (Depping et al., 2016, 2018; Depping & Mandryk, 2017; Halbrook et al., 2019; Harris, 2019; Morschheuser et al., 2017) , communication (Grandi, 2021) and teamwork behaviors including leadership and coordination (Jang & Ryu, 2011; Musick et al., 2021; Williams & Kirschner, 2012). Serious games have been developed as testbeds for teamwork studies, to train teamwork skills (Mayer, 2018; Guenaga et al., 2014; Kutlu et al., 2013; Ellis et al., 2008) foster communication (Grandi, 2021; Handler, 2017) and assess the validity of the environment itself (Mayer, 2018). The next section covers existing work on cooperative video games, investigating design features, patterns, and frameworks. Furthermore, the review covers studies where social interactions have been investigated in cooperative environments. The aim of this review section is to establish cooperative games as environments that can foster teamwork behaviors, and therefore are suitable testbeds for teamwork measurement.

Multiplayer Video Games

Gameplay has been described as the system of interactions between players and game system, and interpersonal interactions between players (Björk & Holopainen 2005; Bergström. et al., 2010). Game structures have been widely described following the Mechanics-Dynamics-Aesthetics (MDA) model (Hunicke et al., 2004). Mechanics are the coded game rules and boundaries. They determine what player actions are supported and how the game environment responds to players' inputs. Dynamics are the resulting interactions between users or the users and the game (Bergström et al., 2010). Aesthetics are the resulting affective responses, such as satisfaction, enjoyment, and frustration.

Interdependence in multiplayer games has been studied to categorize game features. It is defined as the level to which individuals depend on each other, and the nature of this dependence (Depping & Mandryk, 2017; Harris, 2019). Multiplayer games' features can be classified as: "cooperative, competitive, cooperative-competitive, and individualistic" (Morschheuser et al., 2017-p.02), depending on the goal structures (Morschheuser et al., 2017). Individualistic features affect the individual player, such as player power ups. Competitive features support negative interdependence, where players' actions affect each other negatively, such as fighting. Cooperative features support positive interdependence, where player actions affect each other positively, such as sharing resources, and healing each other. And finally, cooperative-competitive features involve team competitions, such as two teams competing to win a game quest. The next section will focus on cooperative video games' design frameworks, patterns, and features.

Cooperative video games: existing patterns

Cooperative video games provide gameplay experiences built on three key characteristics (Morchheuser et al., 2017): cooperative goal structures, mechanics and rules that support cooperation, and communication features. A prominence of shared goal structures has been proposed in a game review conducted by (Zagal et al., 2008). These goals support positive correlations between individual player goals, and therefore provide a team motivation for players to work together (e.g., winning as a team instead of as individual players) (Zagal et al., 2008). Cooperative design patterns are common design features repetitively implemented in multiplayer games to induce player interdependence (Emmerich & Masuch, 2017; Reuter et al., 2014; Zagal et al., 2008). Complementarity is a common design pattern, where players are equipped with game abilities or roles that complement each other's activities (e.g., one player has a shield while the other a sword to fight a common enemy in the game) (Reuter et al., 2014; Seif El-Nasr et al., 2010; Rocha et al., 2006; Zagal et al., 2008) A closely related pattern is synergies between abilities, where players' abilities influence others' abilities positively (e.g., a player using their shield to give another player a jumping boost) (Seif El-Nasr et al., 2010; Rocha et al., 2006; Zagal et al., 2008). Additionally, patterns that encourage players' interactions include abilities that affect other players, including healing abilities and reviving dead player characters (Seif El-Nasr et al., 2010; Rocha et al., 2006; Zagal et al., 2008). Other identified features include shared puzzles, shared characters, and limited life resources (Seif El-Nasr et al., 2010).

Cooperative video games: connection between mechanics and behaviors

Connections between game mechanics and the resulting collaborative dynamics and aesthetics have been suggested. Game mechanics can be implemented to trigger desired outcomes (Mariairs et al., 2011; Peppler et al., 2013; Hämäläinen et al., 2018). For example, complementarity have been suggested to trigger team-members awareness and collaboration through reliance (Wang, 2009; Hämäläinen et al., 2018). Additionally, encrypted information, where players receive unique roles that contribute to the team's mission, has been suggested to encourage collaboration and raise individuals' responsibility toward the team (Hämäläinen & Vähäsantanen, 2011). Furthermore, indirect actions, also referred to as asymmetry in information, can be intended to trigger information exchange to establish a common ground (Collazos et al., 2007; Hämäläinen et al., 2018).

Other mechanics have been suggested to induce cooperative behaviors. Through analyzing World of Warcraft, Left for Dead, Space Alert and Battlestar, Bergström et al., (2010) investigated how they encourage camaraderie. For example, they described that in the board game Space Alert, mutual enemies is a game mechanic used to provide a mutual goal, therefore promoting the dynamics of cooperation and team play. When successful, these dynamics give rise to aesthetics such as team accomplishment and mutual experience. Another example is the mechanic of limited set of actions, which gives rise to the dynamic of communication and coordination, for players to combo their actions and therefore generate stronger attacks. In World of Warcraft, they identified the mechanics of asymmetric abilities and selectable functional roles, where players must ensure that their characters are compatible to achieve the dynamic of team combo. Their work contributed to a variety of new aesthetic patterns, including team strategy identification where teams must determine how to work together. They associated this aesthetic with patterns such as asymmetric abilities, limited set of actions and selectable functional roles. Another aesthetic is team accomplishments, that can arise from simultaneous challenges.

Cooperative video games: asymmetric play

Harris (2019) also followed the MDA framework to suggest a series of mechanics, dynamics and aesthetics associated with asymmetric play. While a mechanic is a rule coded in the game, dynamics arise from the players interactions with the game system (Hunicke et al., 2004). Asymmetry is a cooperative design approach that provides players with different gameplay experiences, through different interfaces, information, roles, and challenges (Harris, 2019; Ouverson & Gilbert, 2021). Therefore, asymmetry generates mandatory interdependence, where players must rely on each other to advance (Harris 2019). In their framework, Harris (2019) suggests a list of asymmetric mechanics, such as asymmetric abilities, where every character fulfills a different role (e.g., medic vs soldier); asymmetric challenges, where players encounter different types of challenges (e.g., cooking a pizza vs delivering a pizza); asymmetric interfaces, for example one player interacting in a 3D environment while the other with a top down 2D view; asymmetric information, such as one player possessing a maze map while the other have to navigate the maze. They also detail the resulting dynamics including unidirectional, where one player depends on the other but not vice versa (e.g., a dead player character waiting for another to revive them); Mirrored dependence where players' dependence on each other is symmetric, for example players shooting a common enemy; Symbiotic dependence, where players' reliance on each other is through different mechanisms, for example one player steering the ship while the other is defending the ship. Additionally, they detail the importance of timing in interdependence, providing categories such as sequential timing (Player A performs their move before player B), and concurrent timing (Player A and B complete their moves at the same time in a continuous manner).

After reviewing the existing body of literature covering cooperative design patterns and frameworks, and how they encourage cooperative behaviors, the next section covers empirical studies where commercial and academic cooperative video games were used to investigate aspects of social play, including teamwork.

Social Play in Cooperative Video Games: Empirical Studies

Cooperative video games can emphasize cooperation and teamwork, through a variety of mechanics and dynamics, described in the previous section. Researchers have been investigating the effects of these environments on a variety of social behaviors, such as ice-breaking, teaming dynamics, prosocial activities, and cooperation. Serious games have been developed to study how cooperative game features can affect group work dynamics. One testbed is Operation Sting where researchers implemented equitable individual contribution, distinctive roles, required social interactions, and concurrent play (Nasir et al., 2015). Moreover, they used designed patterns presented by Seif El Nasr et al., (2011) such as shared puzzles and limited resources. They collected data through video analysis, by counting the words spoken, speech turns, instances and duration of silence, and laughter. They also coded floor holding, where one participant dominates the conversation, and collaborative floor holding where everyone contributes to the discussion. Control (did not play Operation Sting) and Game groups, engaged in a teamwork activity afterwards. They found that groups who played Operation Sting had significantly higher number of words and turns and engaged in collaborative floor holding more than single floor holding, indicating an increase in active participation and frequent alternations in game groups.

To address challenges in virtual teams, such as trust deficiency, low cohesion and group fellowship, and communication difficulties, Ellis et al., (2008) built three different games in Second Life, a 3D multi-user virtual environment: Crossing the Ravine, where players have different pieces with different shapes, once placed together they form a bridge to cross; Tower of Babble, where players have to stack shaped blocks and balance the tower; and Castle Builder where players function as designers and builders, builders cannot see the design and rely on communicated information from the designers. They aimed to design these games to foster social identity, team building and social communication, as proposed solutions for virtual teaming obstacles. To assess whether virtual gaming environment affect team building in organizations, Kutlu et al., (2013) studied Zoom, where some participants received a series of pictures, and must describe them to other participants to place them in the correct order. They examined team aspects through a questionnaire, asking about their team experience, advantages and disadvantages of the exercise, and satisfaction. They compared face-to-face and virtual versions of the game. They reported that participants liked the game environment, avatars, nicknames, and the team building aspects. To understand how cooperative games affect group dynamics and collective intentions, Morchheuser et al., (2017) administered a survey to Ingress game players, a cooperative augmented reality game, where team of gamers must collect virtual objects and open portals. They assessed players' engagement with cooperative features, such as the importance of updating each other's portals, frequency of updating other players' portals and recharging resonators of others. They also measured joint commitment, group norms and anticipated emotions. They found that engaging with cooperative features positively influences the components of group dynamics (e.g., positive emotions, group norms, joint commitment, and team attitudes), which in turn positively affect collective intentions.

To better understand what game properties, foster social ties, Depping et al., (2018), used several measures through a survey, to assess levels of cooperation in games, level of interdependence, and social experiences like relatedness. Their results show that interdependence significantly predicted bonding and bridging social capital. To study the effect of cooperative games in trust formation, Depping et al., (2016) compared a cooperative interdependent game, Labyrinth, where two players play as pusher (can change maze configuration) and collector (can collect coins), and a social icebreaker task. They measured interpersonal trust, and propensity to trust through surveys, and reported that the game was more effective in facilitating trust formation.

CHAPTER 3. APPROACH

Study Overview

This study aims to develop and apply a teamwork measurement system in cooperative gaming environments, in addition to exploring how cooperative design features induce teaming behaviors. The study was conducted through annotating gameplay footage of cooperative video games, publicly accessible on YouTube. The use of existing cooperative games, played in naturalistic settings, was followed to study player interactions emerging authentically, in complex and fully developed video games, therefore attempting to gain a more authentic representation of how teams of gamers naturally interact in the gaming environments (Harris, 2019; Isbister, 2010). The approach section details the process of developing the codebook, selecting video games, and the data analysis plan.

A codebook was developed to guide the annotation process. A codebook is a compilation of tags or labels, referred to as codes, which assign segments of interpretations to the textual or audio-visual data (Miles & Huberman, 1994). The codebook constituted of a list of teamwork behavioral markers and cooperative games' design features. These codes were used to label the gameplay footage whenever an observable event was detected. Coders labeled every teaming event by annotating a behavioral marker with the cooperative feature associated with it in the game. Cooperative gameplay videos were selected following cooperative and technical criteria.

The gameplay coding process generated quantified behavioral markers. Additionally, it extracted empirical data on the frequency of associations between cooperative features and teaming behaviors. Therefore, using the collected data, the data analysis plan details how every research question will be addressed to explore the consistency of cooperative video games in inducing teaming behaviors and how they are associated with cooperative features. Additionally, performance-based comparisons were conducted to gain insights on how teaming behaviors affect performance outcomes. The methodology section will cover the process followed to collect and analyze the data and will be divided into three major sections: the codebook development, the video game selection, and the data analysis plan.

Methodology Motivation

Coding gameplay footage to track behaviors and design characteristics has been employed in previous work. Awareness cues in cooperative games were analyzed, through a grounded theory approach, by watching gameplay videos from Youtube and Twitch, conducting open coding to identify awareness cues, and then conducting axial coding to structure the codes (Wuertz et al., 2018). Grounded theory was also used by Toups et al., (2014) to identify cooperative communication mechanics, through analyzing Twitch gameplay footage and identifying game mechanics. In these two applications, grounded theory was an approach used to derive frameworks from collecting raw data through observations. Cooperative design patterns were studied in a variety of cooperative video games, where footage was analyzed to extract common design features, through a codebook of design mechanics and cooperative behaviors (Seif El-Nasr et al., 2010). They annotated gameplay footage to track cooperative features and associated them with the resulting behaviors, using commercial cooperative video games. Coordination behaviors in Warcraft were studied through visual and auditory gameplay recordings to analyze players behaviors (Williams & Kirschner, 2012). To better understand how asymmetric mechanics affect social behaviors, Harris (2019) conducted thematic analysis on gameplay footage of their asymmetric game Beam Me Round Scotty and reported observations. Therefore, previous work supports the suitability of using a gameplay observation approach to identify game mechanics and behaviors.

Methodology Process

Figure 1 provides a visual representation of the process conducted to measure teamwork in cooperative gaming environments. The process consists of two major parts: codebook creation and testing, and video game and footage selection.

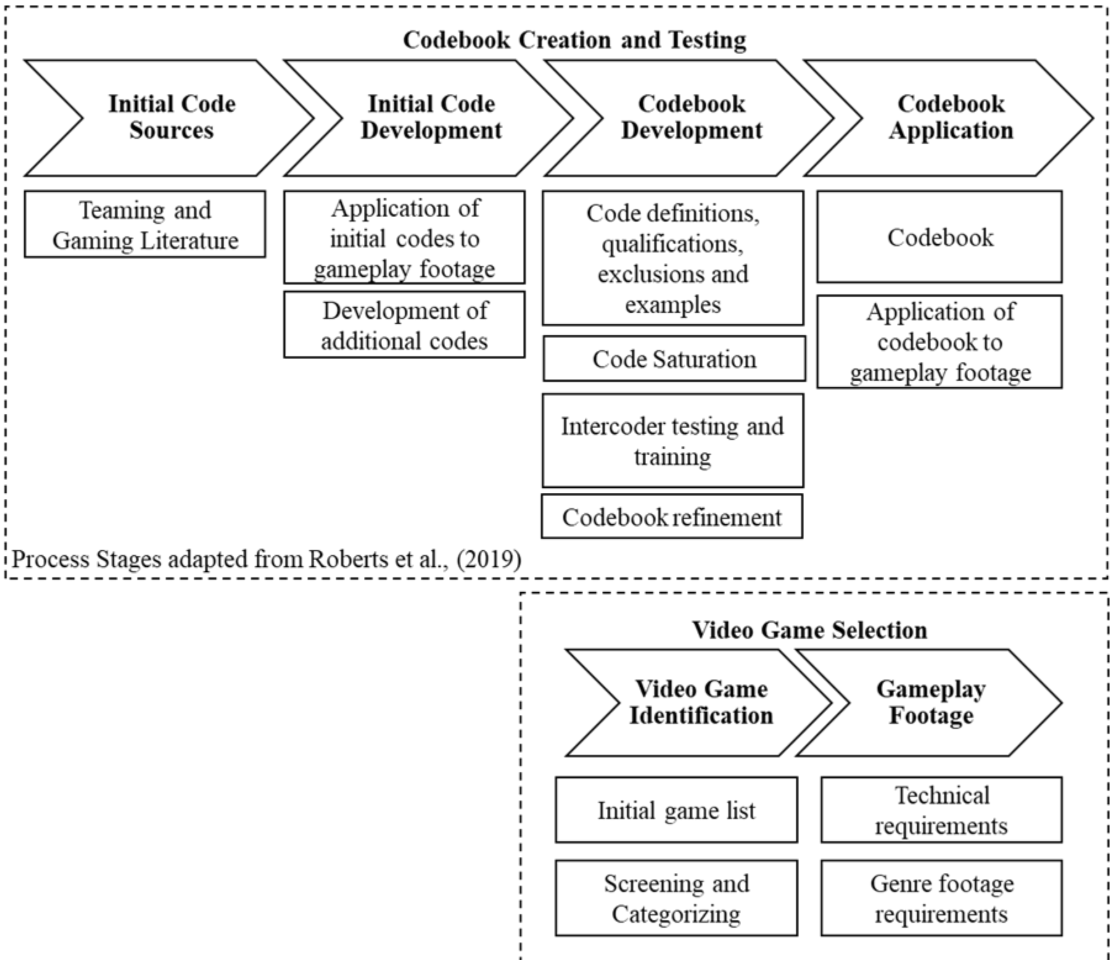


Figure 1. Codebook development and video game selection process

Codebook Creation and Testing

A codebook was developed to annotate the gameplay footage, using codes to label the teaming events associated with cooperative features. The process consisted of four steps: initial code sources, initial code developments, codebook development, and codebook application. The codebook in this work is a compilation of labels, used to assign a unit of meaning Huberman

(1994), to observable events happening in the gameplay footage. The unit of meaning in this context is a behavioral marker associated with a cooperative feature.

Initial Code Sources

Codes can either be theory driven, compiled from existing literature and theory, or data-driven, derived from raw-data (DeCuir et al., 2011). The first iteration of the codebook development was theory driven. The researchers compiled a list of initial code sources from previous teaming and gaming literature.

Five teaming competencies were suggested by Salas et al., (2005) to achieve teaming effectiveness: Team Leadership, Mutual performance monitoring, Team orientation, Backup Behavior and Adaptability. Using these five competencies as a guide, a list of relevant behavioral markers was compiled from existing work (Marks et al., 2001; Rosen et al., 2011; Rousseau et al., 2006) . An existing list of behavioral markers was provided in an adaptive teamwork framework with the following stages: Situation Assessment, Plan Formulation, Plan Execution and Team Learning (Rosen et al., 2011). The framework describes the teamwork processes as a series of teaming stages, where each stage has an associated list of behavioral markers. Based on previous work, a theory-driven list of teaming competencies was compiled:

1. Planning Phase: Situation Assessment, Mission Analysis, Strategy Formulation, Team Leadership
2. Coordination: Implicit Coordination, Explicit Coordination
3. Monitoring: Mutual Performance Monitoring, Backup Behavior, Systems Monitoring
4. Adaptability: Contingency Planning, Reactive Strategy, Behavioral Adaptability and Team Learning
5. Cohesion and Social: Interpersonal Relationships, Task Cohesion

An initial list of cooperative features was developed based on literature. Rocha et al., (2008) provided cooperative design patterns: complementarity, synergy between abilities, abilities that can affect other players, shared goals, and special rules that enforce cooperation. This list was expanded on by Seif El-Nasr et al., (2010) who further analyzed a variety of cooperative video games and suggested additional features: same object interactions, shared puzzles, limited resources, and shared characters.

Initial codebook development

Application of initial codes to gameplay footage

The codebook development followed a mix of deductive and inductive approaches. Deductive approaches assume that existing principles can be applied to the studied phenomenon (Fereday et al., 2006). Therefore, following a deductive approach, the preliminary list of codes, developed from existing principles and sources, was applied to gameplay footage of three video games: It Takes Two (Hazelight Studios, 2021), Trine 4: The Nightmare Prince (Fronzebyte, 2019) and Don’t Starve Together (Klei Entertainment, 2016). It Takes Two and Trine 4 are puzzle platformer video games, with mandatory teamwork. The initial approach and preliminary results of these codes is demonstrated in Farah et al., (2022) where six teams were coded, two in every video game, to compare frequency of teamwork behaviors. This work expands on the preliminary work by applying a final version of the codebook to 18 video games categorized into four genres, therefore aiming to establish a more generalizable approach to testbed design through teamwork assessment in cooperative games. Mandatory teamwork implies that the design of the game forces the players to work together to proceed. While Don’t Starve Together is a survival game with emergent teamwork. This implies that players can choose different ways to navigate the game, and they are not forced to engage in teamwork on a game mechanical level. Therefore, these games provided a mix of mandatory and emergent teaming situations, allowing researchers to test the initial codes in different contexts and observe new behaviors and features. While watching the gameplay footage of six teams (two teams within every video game), one researcher labeled the behaviors and features following the initial list. Additionally, notes about the behaviors were taken and new cooperative features and behavioral markers were added when noticed through an inductive approach.

Development of additional codes

After applying the existing codes to the gameplay footage, the coder took notes of teamwork behaviors and cooperative features that were not initially in the list. This allowed the researcher to gain a better understanding of how the behaviors and features appear in the games, and to add new sub-categories. For example, while the initial list included “shared obstacles”, the coder noticed that in the games shared challenges can be separated into shared obstacles and shared puzzles. Obstacles were challenges that only need one set of actions to be cleared, while puzzles required a series of interdependent tasks to be cleared. Therefore, through an inductive approach, additional cooperative features were added.

Codebook development

The next section details the steps of developing the codebook after the initial sourcing and initial code development.

The codebook was divided into categories, following teamwork processes suggested in Marks et al., (2001). The transition processes included situation assessment, mission analysis, plan formulation and team leadership. The action processes included explicit coordination, implicit coordination, mutual performance monitoring, backup behavior and systems monitoring. The adaptive processes included reactive strategy adjustment and team learning-reviewing. Finally, the interpersonal processes included team cohesion and social interactions. Within every process, teaming competencies were first defined, followed by a table of behavioral markers.

Every behavioral marker had a definition, description in games, qualifications, examples, and codes. This approach follows the guidelines originally provided by Boyatzis (1998) and is followed in code development in qualitative research (Roberts et al., 2019).

The definition of the behavioral marker was based on existing definitions from teaming literature, from which the behavioral markers were originally extracted. The description in games aims to establish a link between the literature definition and how the behavior applies in the gaming context. The qualifications aim to detail what would qualify an event in this behavioral marker. Examples include some general examples extracted from annotated games to demonstrate the code application. And finally, the code includes the acronym for the behavioral marker.

Codebook iterations

To develop the codebook, every teaming competency was divided into behavioral markers, with definitions, descriptions, qualifications, examples, and an acronym. As for the cooperative features, every feature had a definition and an acronym. To refine the code definitions and test their applicability in different gaming contexts, four coders used the codebook in seven games from different genres: It Takes Two (Puzzle Platformer), Portal 2 (Puzzle Platformer), Don’t Starve Together (Survival), Keep Talking and Nobody Explodes (Asymmetric), Lovers in a Dangerous Spacetime (Simulation), Overcooked 2! (Simulation). Throughout the process, the codebook definitions were refined through testing them in different gaming environments by four coders. To iterate the codebook, the four coders would code footage from the games and meet to discuss the codebook application and definitions of behavioral markers and cooperative features. Adjustments were made to refine the definitions, develop further codes, and group codes.

Code saturation

In the codebook development phase, code saturation is achieved when it is perceived that a sufficient representation of the theme (in this study, the teamwork construct) is accomplished through the codes (Roberts et al., 2019). For example, if situation assessment is a construct in the codebook, the code saturation is achieved when the behavioral markers to be coded under this construct are established, and they are deemed as sufficient demonstration of the construct. Once this saturation is established, it implies that no further behavioral markers will be represented by new codes for this construct. To achieve code saturation, the coders would code together as a group, to assess the behaviors and discuss the codebook's definitions and how they apply to the gaming context. The coders would also code independently and then compare codes to find similarities and differences. When suggesting new behavioral markers, the coders would discuss whether they fit under an existing category or if it is a new behavioral marker. Same for cooperative features. After coding in different gaming environments and genres, the coders were not finding new behavioral markers or cooperative features to add to the codebook, therefore perceiving a code saturation after four iterations. The first iteration of the codebook included 19 codes of cooperative features, compared to 32 codes in the final iteration. Several cooperative features were added after testing the codebook with different genres, such as environment modifying ability (EMA), extracted from Portal 2's opening portals ability, and moving object (MOB) resources (MR). Furthermore, cooperative features initially were not grouped into categories. In the final codebook they were grouped into five categories: gameplay features, player abilities, environmental components, story aesthetics and game world, and gameplay dynamics. While for behavioral markers the first version included 30 codes and was brought down to 23 after the four iterations. Overlaps were noticed between some behavioral markers and therefore they were eliminated. For example, team leadership was initially broken down into four behavioral markers, but after iterating they were deemed overlapping and hard to distinguish, and were grouped into one team leadership code, that addresses the main functions of leadership. Implicit performance monitoring was also eliminated since no common ground was established to track performance monitoring that does not happen verbally or behaviorally.

Inter-coder testing and training

The fifth version of the codebook was tested through an intercoder agreement process to assess the consistency of judgment that can be achieved when different raters use the codebook (Roberts et al., 2019). By implementing inter-rater agreement in the development phase, the process can guide the development to establish definitions of behavioral markers that can be judged consistently by multiple raters. Furthermore, in observational teamwork studies, inter-rater agreement is recommended to mitigate subjectivity and bias, and to establish specific behavioral markers with clear coding guidance (Salas et al., 2017). The inter-coding agreement process aimed to refine the final coding process with assistance of another coder, and to establish a coding perspective that relies on objective definitions agreed on by two coders rather than one coder.

The two coders were assigned footage from a selected video game, ranging between 10 to 15 minutes. In the first iteration, the two coders would first code independently, and then meet to discuss agreements and disagreements. An inter-coder agreement would be generated. The coders would code the same footage again to verify enhancements resulting from discussions. Afterwards, new footage would be assigned. The same process would be repeated (code independently, meet and discuss, and then code again). The process was repeated a maximum of three times, if the inter-rater agreement is below 75%, a minimum percentage recommended to demonstrate adequate agreement. By the end of the third iteration, the highest inter-rater agreement was chosen, even if it was below 75%, and the notes and discussions associated with that round would be used to adjust the codebook's definitions. The iterations aimed to introduce the video game and develop an understanding of how the codebook is applied in the gaming environment. After the coders became familiar with the game cooperative features and teamwork behaviors, the second iteration aimed to build a shared understanding of how the behaviors should be coded. Finally, the third iteration assessed the overall enhancement in the coding process and provided the final coder with a final understanding of how to code the behaviors. The games selected to go through this process covered all four genres.

To calculate the inter-rater agreement percentage, codes from two raters for the same footage were sectioned into puzzles (if the puzzle is equivalent to the level) or time chunks ranging from one to two minutes.

Figure 2 and Figure 3 demonstrate the process in Keep Talking and Nobody Explodes (Steel Crate Games, 2005) and It Takes Two (Hazelight Studios, 2021). The total number of every behavioral marker coded by every coder is calculated within the puzzle or the time section. The number of agreements is calculated by deriving the minimum (for example for SA-CR, both coders agreed seven times). Matching behavioral markers would be considered an agreement if they fall within the sectioned time, in this case the symbols and the wires puzzle in keep talking (approximately two minutes). The total codes were calculated by obtaining the maximum between the two coders (For example SA-CR was coded eight times). The percentage agreement was calculated by dividing the total agreement by the total codes (minimum over maximum).

	Coder 1				Coder 2		
Symbols	1:01	SA-CR	SA-CM		0:57	SA-CR	SA-CM
	1:12	SA-CR			0:59		
	1:18	SA-CR			1:16	SA-CR	
	1:27	MA			1:26	SA-CR	
	1:32	SA-CR			1:30	MA	
	1:42	SA-CR	SA-CM		1:40	SA-CR	
	1:43	SA-CM			1:40	SA-CR	SA-CM
	1:46	SA-CR			1:55	SA-CM	
	1:56	SA-CR	MA		2:00	MA	SA-CR
	1:56	EC-R			2:31	EC-R	
	2:05	MA			2:41	SA-CR	SA-CM
	2:35	SA-CR	SA-CM		2:52	EC-S	TC-E
	2:47	EC-S	TC-E		2:52	SA-CR	
Wires	2:50	SA-CR	SA-CM		2:57	SA-CM	
	2:57	SA-CR	SA-CM		3:06	SA-CR	SA-CM
	3:06	SA-CM			3:11	SA-CM	
	3:12	EC-R			3:16	EC-R	
	3:13	SM-E			3:22	SM-E	
	3:19	EC-R			3:42	EC-R	
	3:26	MA			3:49	EC-R	
	3:45	EC-R			4:07	MPM-E	
	3:50	MPM-E			4:10	EC-R	
	3:53	SA-CR			4:10	EC-S	
	4:09	EC-S			4:18	TSKF	TMF
	4:10	TSKF	TMF		4:38	TL-R	

Puzzle		Coder 1	Coder 2	Agreemen	Total	% Agreement
Symbols	SA-CR	8	7	7	8	
	SA-CM	4	4	4	4	
	MA	3	2	2	3	
	EC-R	1	1	1	1	
	EC-S	1	1	1	1	
	TC-E	1	1	1	1	
	Total	**18**	**16**	**16**	**18**	**89%**
Wires	SA-CR	3	2	2	3	
	SA-CM	3	3	3	3	
	EC-R	3	4	3	4	
	SM-E	1	1	1	1	
	MA	1	0	0	1	
	MPM-E	1	1	1	1	
	EC-S	1	1	1	1	
	TSKF	1	1	1	1	
	TMF	1	1	1	1	
	TL-R	0	1	0	1	
	Total	**15**	**15**	**13**	**17**	**76%**

Figure 2. Inter-coder agreement example from Keep Talking and Nobody Explodes

Coder 1				Coder 2			
33:45:00	SA-CR	SA-CM		33:47:00	SA-CR	SA-CM	
33:53:00	IC-S			33:52:00	MA		
33:57:00	EC-R			34:00:00	SA-CR	SA-CM	
34:07:00	IC-S			34:09:00	IC-S		
34:19:00	IC-S			34:20:00	IC-S	TC-E	
34:22:00	TC-E			34:35:00	EC-S		
34:35:00	MPM-E	EC-S		34:40:00	EC-S		
34:39:00	EC-S			34:42:00	TSKF		
34:43:00	TSKF			34:52:00	TSKF		
34:52:00	TSKF			35:06:00	TSKF		
35:05:00	TSKF			35:14:00	SA-CR	SA-CM	
35:13:00	SA-CR	SA-CM		35:21:00	EC-S		
35:21:00	SM-I			35:27:00	SA-CR	IC-S	
35:29:00	SA-CR	SA-CM		35:39:00	TC-E		
35:31:00	IC-S			35:49:00	IC-A		
35:35:00	TC-E			36:00:00	IC-S		
35:52:00	IC-A						
36:00:00	IC-S						

Time Section		Coder 1	Coder 2	Agreemen	Total	% Agreement
Section 1	SA-CR	1	2	1	2	
	SA-CM	1	2	1	2	
	IC-S	3	2	2	3	
	EC-R	1	0	0	1	
	TC-E	1	1	1	1	
	MPM-E	1	0	0	1	
	TSKF	2	2	2	2	
	MA	0	1	0	1	
	EC-S	2	2	2	2	
	Total	12	12	9	15	60%
Section 2	TSKF	1	1	1	1	
	SA-CR	2	2	2	2	
	SA-CM	2	1	1	2	
	SM-I	1	0	0	1	
	IC-S	2	2	2	2	
	TC-E	1	1	1	1	
	IC-A	1	1	1	1	
	EC-S	0	1	0	1	
	Total	10	9	8	11	73%

Figure 3. Inter-coder agreement example from It Takes Two

Several limitations apply to this method. First, it was assumed that the maximum between the two coders entails the total number of codes for the addressed behavioral marker. Second, this method does not account for chance. Third, it does not account for missing data. However, percentage agreement is still reported in literature, and can serve as a simple guide on the agreement between two raters, that can account for a high number of variables, which is convenient in this use case (Roberts et al., 2019).

Table 1 summarizes the inter-rater agreement values. To assess the codebook on all genres, two to three games from every genre were tested. Time of videos assigned for inter-rater agreement ranged from 10 to 20 minutes. The inter-coder values range between 63% and 78%, which indicate moderate levels of agreements.

Table 1. Inter-rater agreement values

Video Game	**Sections**	**Genre**	**Agreement**
Keep Talking and Nobody Explodes	Puzzles	Asymmetric	68%
KeyWe	Puzzles	Simulation	72%
Portal 2	Puzzles	Platformer	77%
Lovers in a Dangerous Spacetime	One-two minutes	Simulation	73%
Don’t Starve Together	One-two minutes	Survival	78%
Overcooked 2	Puzzles	Simulation	73%
It Takes Two	One-two minutes	Platformer	69%
The Survivalists	One-two minutes	Survival	63%
We Were Here Forever	One-two minutes	Asymmetric	63%

Application of codebook

Codebook design: behavioral markers

This section presents examples of the final codebook structure. The full codebook is attached in Appendix A. The behavioral marker definition is from existing literature, the description in games describe how the behavioral marker applies in gaming environments, the qualifications describe technical characteristics of the behavior that qualifies it to be coded as the behavioral marker, and the code presents the acronym to be used in the coding process.

Table 2 shows an example from situation assessment-cue recognition. Cue recognition would be observed as the character or avatar scanning the environment, through observable actions (e.g., looking around, scrolling the screen) and detecting a cue by verbally expressing it (e.g., There is a door there, I see a button next to the staircase). When the team member signals that they have detected a cue, the behavior is coded as SA-CR.

Table 2. Situation assessment-cue recognition behavioral marker from the codebook

Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Cue recognition	A team member (or more) scanning the environment for cues that can influence the mission. (Marks et al., 2001; Rosen et al., 2011)	A team member's avatar is visibly seen scanning the environment (looking around, scrolling the screen) as an initial situation assessment to detect cues. Verbal expressions give signs that the member is scanning or looking for cues, such as relaying information that they see cues ("I see a button")	-To be behaviorally seen or verbally heard -It must be in the transition process (hence, it is coded when the team engage in this scanning when they first encounter the task or the mission and still figuring out how to proceed)	-"There's a door there" -"I see a button next to the staircase"	SA-CR

Table 3 demonstrates another example of explicit coordination sequencing/synchronizing. The description in games specifies that this is a verbal behavior of sequencing tasks through assigning a sequence of roles or synchronizing movements. For example, "1,2,3, go"-synchronizing a two-button jumping action or "I put my leg first and then you go next"-BiPed color changing platform.

Table 3 Explicit coordination sequencing or synchronizing behavioral marker from the codebook

Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Synchronizing or Sequencing	Team members pacing their activities through verbally sequencing tasks or synchronizing movements.	Sequencing task roles in a certain order to execute interdependent game activities. Team members synchronize when their movements through timing. -Player A providing information to sequence Player B's actions -Player A and B provide information to sequence each other's actions	- Verbally heard -Sequencing or synchronizing here is an action process and therefore it's not coded when it's a part of the plan formulation.	-"1,2,3, go!" -"I'll get the wood first and you can add the grass"	EC-S

In contrast, Table 4 demonstrates an implicit coordination example, where players sequence or synchronize their actions without verbal communication. The examples provide demonstrations from Overcooked 2 when players assemble the dish without verbally sequencing the steps.

Table 4. Implicit coordination sequencing or synchronizing behavioral marker from the codebook

Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Sequencing or synchronizing	When Team-mates sequence or synchronize their actions (Marks et al., 2001) (same functions as explicit coordination) however without explicitly communicating	In a sequential task, when players do a game task in sequence or synchrony without communicating (Wuertz et al., 2018), or when synchronizing actions (e.g., jumping on buttons), they do it without communicating explicitly.	-It must be behaviorally observable (therefore an action or a movement by the player avatar or character)	-Team mates jumping on two buttons at the same time without communicating -Team mates ordering their ingredients (e.g., Overcooked) without communicating	IC-S

Codebook design: cooperative features

Table 5 summarizes some cooperative features from the codebook.

Table 5. Examples and brief descriptions of cooperative features in the codebook

Cooperative Feature	**Brief Description**	**Example**	**Code**
Complementary Puzzle	A game challenge that requires a series of actions to be cleared. Players are equipped with complementary abilities.	Hammer and Nail puzzle in It Takes Two.	CP
Task Allocation Continuous Puzzle	A continuous puzzle, where players are continuously working on a common task.	Cooking level in Overcooked 2!	TACP
Environment Modifying Abilities	Players equipped with abilities to modify the environment (e.g., opening portals)	Ability to open portals in Portal 2.	EMA
Crafting Abilities	Players equipped with abilities to craft new utilities using resources.	Ability to craft science machines in Don’t Starve.	CA
Environmental Resources	Resources that can be collected by players.	Grass, food, gold in survival games.	ER
Common Risks	Risks that impose danger on more than one player and affect their life status.	Darkness that causes insanity in Don’t Starve Together.	CR

The cooperative features section of the codebook divided the features into five categories: gameplay features, sharing abilities, environmental components, story aesthetics and game world, and gameplay dynamics. The full list of categories and cooperative features can be found in Appendix A.

Coding rules

The codebook includes coding rules to enhance the consistency of coding. First, up to three behaviors can be grouped into one annotation if they are happening simultaneously or sequentially within a minute of time. Furthermore, the same behavior cannot be labeled with two codes. However, behaviors can happen simultaneously (for example monitoring the environment and reporting what is being observed; or observing a teammate' performance and providing feedback). Behaviors can be cognitive (e.g., observation), action-based (e.g., moving character, moving objects) or verbal (e.g., verbally speaking). Therefore, a behavior in the same category (e.g., cognitive), should not be labeled with two codes. Finally, up to two cooperative features can be paired in one annotation if they occur simultaneously with the induced behavior. For example, a player looks around the environment to find resources. The cooperative features associated with the action are SE/ER (shared environment, environmental resources). Both features were associated with the monitoring behavior.

Application of codebook to gameplay footage

The author proceeded in coding the selected gameplay footage. The final round of coding took place over four months. The coder attempted to code one full genre at a time, finishing one video game within a genre with all its teams before moving to the next one. This approach was designed to achieve several coding benefits.

First, to ensure consistency of coding within a genre, all the games in a genre were coded before moving on the next genre. Behavioral markers can have different applications depending on the environment and need to be adapted when applying them to new contexts (Grand et al., 2013). The coder aimed to develop a clear mental model of how the behaviors occur in one genre, and code the full genre before moving to the next. Coding one genre proved to be more time efficient for the coder, since switching back and forth between different genres would slow the coding process for the coder to readapt to the genre approach.

Video Game Selection

The game selection process started with identifying a pool of video games retrieved from internet recommendations, following search keywords. The video games were screened following cooperative and technical criteria. Cooperative criteria aimed to assess whether the game is cooperative. Technical criteria aimed to assess the gameplay footage for annotating purposes. Games were categorized into four genres. The process was concluded after selecting four to five games within every genre. This section details the video game selection stages.

Genre categorization

The genres were categorized as follows: Puzzle Platformer, Simulation, Asymmetric and Survival. In this study, the research question aims to identify how video games within genres and between genres induce teaming behaviors. Therefore, the genre definition focused on the interactive aspects of games, rather than the aesthetics or artistic aspects (Apperley, 2006). Puzzle Platformer genre included platforming games with a focus on puzzle solving (Apperley, 2006; Tietojenkäsittely, 2016; Vargas-Iglesias, 2020). Simulation genre grouped the games that were simulating the same activity repeatedly, relying on one core game loop in every game level (Apperley, 2006; Sicart, 2015). Asymmetric genre had games with a total asymmetry throughout the whole gameplay, with players being separated by interface, information, and roles (Harris, 2019). And survival genre included games with crafting survival elements (Sicart, 2015).

Hence, the genre classification was guided by how the games were intended to be played. Table 6 breaks down the differences between the genres, following a five-mechanic categorization of physics, economy, progression, tactical maneuvering, and team interactions (Adams & Dormans, 2012). The table is adapted from Adams & Dormans, (2012), and applied to the four genres analyzed in this study.

Table 6. Genre categorization over five game mechanics

	Physics	**Economy**	**Progression**	**Tactical Maneuvering**	**Team Interactions**
Puzzle Platformer	Detailed physics (moving, jumping, shooting)	Power ups and collectables in the gaming environment	Predesigned levels. Tightly controlled solutions.	Puzzle Solving	Induced by mandatory interdependent tasks and shared goals.
Asymmetric	Simple physics. Detailed physics if 3D hosted.	Power ups and collectables in the gaming environment	Predesigned levels. Tightly controlled solutions.	Puzzle Solving	Induced by mandatory interdependent tasks and shared goals.
Simulation	Simple physics (for character movement)	Environmental resources, collectables, and power ups	Predesigned levels. Loosely controlled solutions.	Task Allocation.	Induced by emergent and mandatory interdependent tasks and shared goals.
Survival	Simple physics (for character movement)	Environmental resources (food, resources), hunting requirements, inventories, and crafting.	Generated maps with different layouts. Loosely designed and controlled.	Team tactics. Resource management and economy building.	Induced by emergent interdependence to support survival demands.

PRISMA and Included Video Games

Figure 4 presents the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) diagram (Moher et al., 2009), detailing the process of compiling the initial game list, cooperative screening, and technical screening.

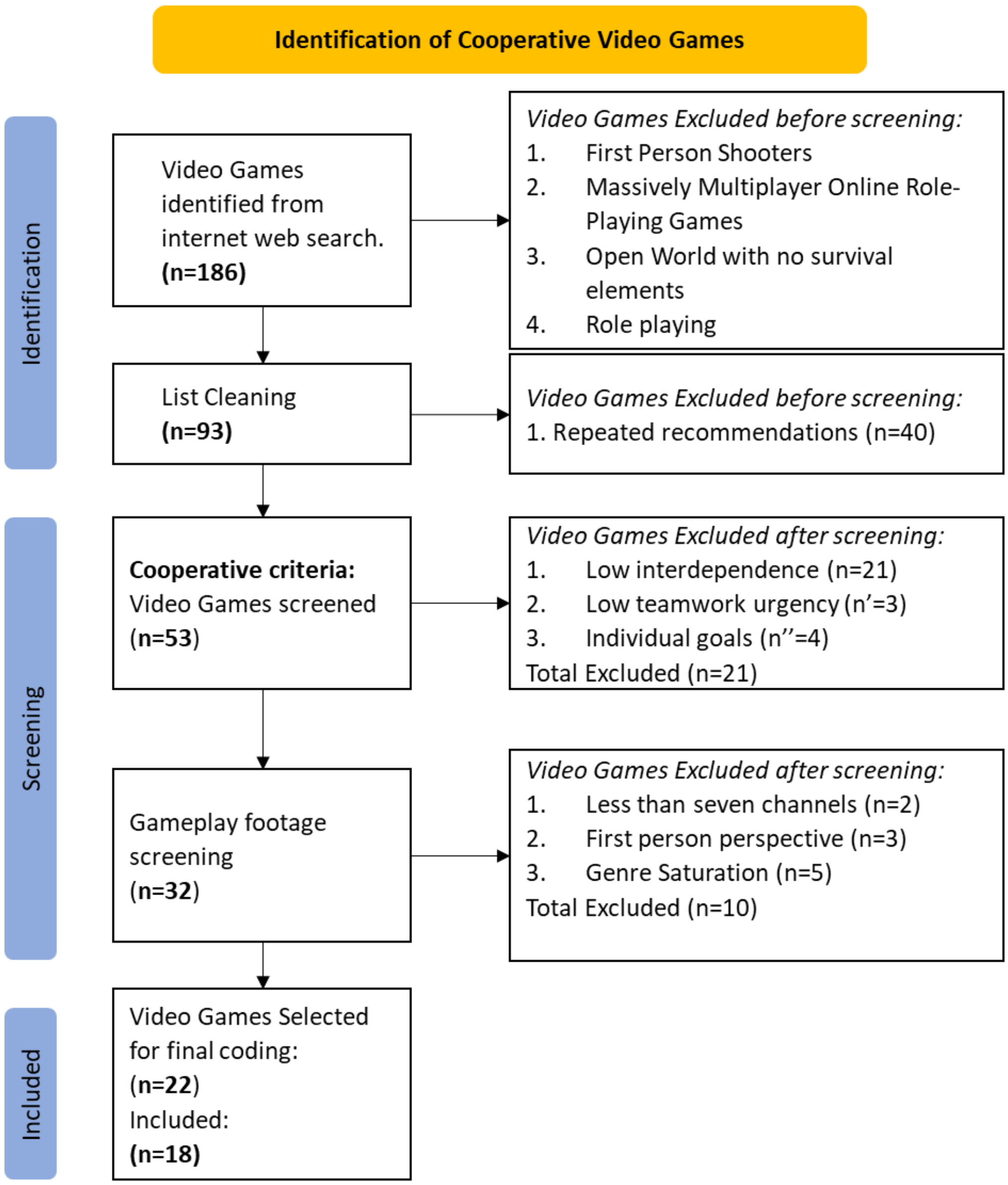


Figure 4. PRISMA diagram for video game identification and selection

The PRISMA guidelines start with choosing eligibility criteria (inclusion and exclusion criteria) and explaining how the studied records (in this case video games) were grouped. In this context, the eligibility criteria were divided into cooperative and technical criteria. Cooperative criteria aim to assess whether the games support cooperative goals and mechanics, while technical criteria aim to assess whether the gameplay footage is eligible for annotating (e.g., clear commentary, footage quality). Afterwards, the PRISMA involved identifying a search strategy and information sources. In this study, the search strategy involved internet search keywords to identify recommended lists of video games from websites. The following paragraphs detail the PRISMA process stages.

Identification

The process started with collecting cooperative video games through internet searches using keywords and referring to gaming websites and YouTube recommendations. Keywords were used to screen recommended cooperative video games on a variety of websites through google search. Table 7 summarizes the list of keywords and corresponding websites. After conducting an internet search, using the keywords, the author scanned the top search results, with titles matching the keywords. Title examples include “Best co-op games to play right now with friends and family” (GamesRadar-Loveridge 2022), “The 17 Best Co-op Games of all time” (Vg247-Raynor, 2023), “25 Best Online Co-Op Games” (Wired- Hill, 2022). Table 7 details the keywords used and the resulting websites from which the initial pool of video games was identified.

Table 7. Web search keywords and websites

Key Word	Website
Cooperative Video Games	"The Best co-op games to play right now with friends and family"- (Games radar- Loveridge, 2022)
	"The Best co-op games of all time"- (Vg247-Raynor, 2023)
	"35 amazing addictive couch co-op games"- (Wired- Hill, 2022)
	"The best co-op games to play after beating It Takes Two"- (Gamerant-Gingerich, 2023)
	"The 10 best co-op video games"- (WatchMojo, 2022)
Puzzle Platformer co-op video games	"13 co-op games for gamers that love puzzles" (Gamerant-Stalberg, 2023)
Multitasking co-op video games	"20 great multitasking games similar to overcooked"-(Gamerant-Lagiola, 2023)
Asymmetric video games	"The best asymmetrical multiplayer video games"- (The Gamer-Alston, 2022)
Multiplayer survival games	"18 multiplayer games to play if you like Don't Starve Together"-(Gamerant-Kurland, 2022)
Multiplayer cooperative games	Google search recommendations- from sources across the web.

To include a variety of game genres in the initial pool of games, more specific key words were added including puzzle platformer cooperative games, puzzle cooperative video games, survival multiplayer video games, and asymmetric video games. Following recommendations from Webpages and YouTube videos, researchers compiled a list of 53 initial video games for review, excluding first person shooters, massively multiplayer online role-playing games, and open world with no survival elements. The full list along with the screening process is provided in Appendix B.

Screening

Researchers started to review the compiled list through watching 5-10 minutes of gameplay footage. For a game to be selected to proceed in the screening process, it had to meet the following cooperative criteria, to determine whether the game supports cooperative goals and mechanics. The criteria were proposed by Morchheuser et al., (2017):

- The game is cooperative: supports shared goals
- The game has rules and mechanics that support interdependence
- The game supports verbal communication channels

Table 8 shows an extract of the table where the games were compiled, along with the tracked characteristics to make the selection. The full screening table is presented in Appendix B.

Table 8. Example of the video game screening process

	Game	**Genre**	**Shared Goals**	**Verbal Communication**	**Interdependence**	**Couplin g**	**PASS/ FAIL**
1	It Takes Two	Action Puzzle Platformer	Yes	Yes	High	Closely	PASS
2	Don't Starve Together	Survival	Yes	Yes	Intermediate	Loosely	PASS
3	Keep Talking and Nobody Explodes	Asymmetric al	Yes	Yes	High	Closely	PASS
4	Lovers in a dangerous spacetime	Space Shooter	Yes	Yes	Intermediate	Closely	PASS

The table covers the specific genre of the game assigned on the game developer's webpage, and assesses whether the game has a shared goal, verbal communication, the level of interdependence and whether it is closely or loosely coupled.

Interdependence levels

Interdependence is the level of positive reliance players must engage in to progress in the game. Harris (2019) suggests that interdependence levels can be manipulated and identified as high, intermediate, or low.

For this work, high interdependence was assigned to games specifically designed for cooperative experiences with mandatory interdependence. Mandatory interdependence is at a mechanical design level (Harris, 2019).

Intermediate interdependence was assigned to games that support multiplayer experiences through providing players with environments of mandatory and optional interdependence. Players rely on each other to achieve the shared goal but can execute a variety of tasks by themselves without needing other players. For example, the game Don't Starve Together support optional interdependence where players can share resources with each other, craft common utilities and fight enemies together, and mandatory interdependence where players can only be revived by other players.

Low interdependence was assigned to games that lack rules and mechanics of mandatory interdependence. Low interdependence is an example of games where a co-op mode was included in originally single player campaigns, with little to no alterations to the game design to support cooperative actions, therefore players can play with each other in the environment, without needing to work together (Wendel & Konert, 2016). Interdependence criteria is summarized in Table 9.

Table 9. Interdependence level criteria

Interdependence Level	**Criteria**
High	**Mandatory interdependence** dominates the game. Designed at a **mechanical level** where players need each other to clear the game levels.
Intermediate	A mix of **mandatory** and **optional interdependence**. The game supports optional interdependence through sequential tasks or game pressures, including time limits, overload, threats, and survival elements.
Low	The game has a shared goal (cooperative). Mandatory interdependence is **rarely** implemented, and mechanics and rules **do not impose urgency** for optional interdependence.

Coupling

Games can be characterized as closely or loosely coupled. In closely coupled games, player characters are forced to maintain close distance and cannot do their tasks before having to interact with other teammates (Tang et al., 2006). For example, the game level stops advancing for one player until the second player reaches the same level. Alternatively, loosely coupled video games allow player characters to advance and explore the game by themselves and to execute their tasks for relatively longer periods (compared to close coupling) without having to maintain a close game distance nor mandatory interactions with other players (Tang et al., 2006).

Game selection thresholds

Figure 5 represents the final game selection process, following an assessment of the three cooperative criteria: shared goals, verbal communication, and interdependence level.

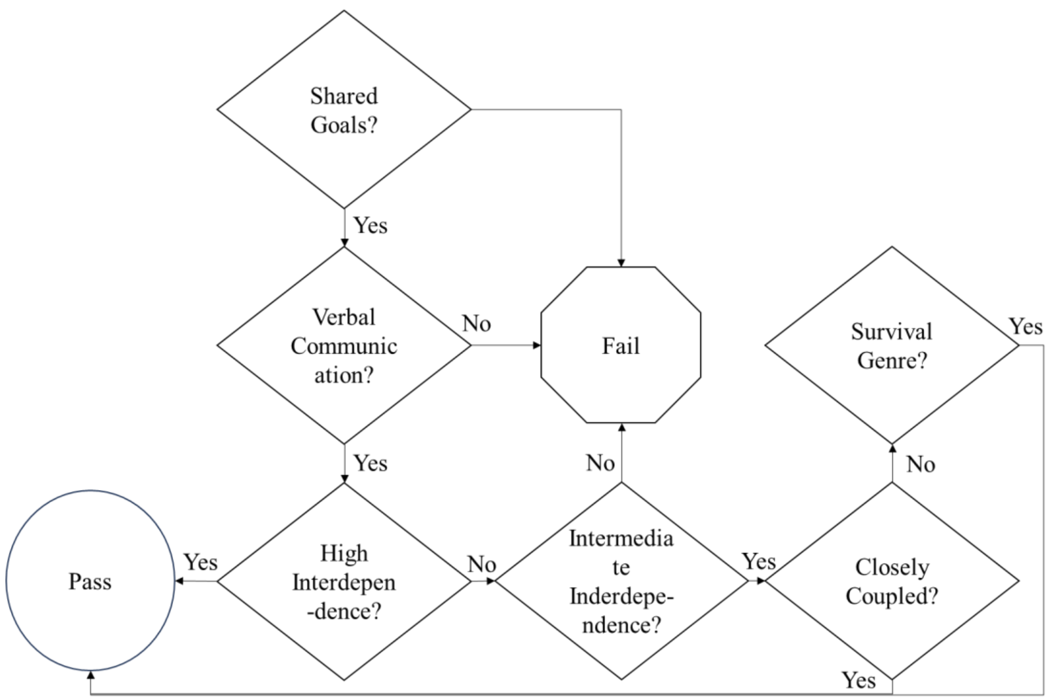


Figure 5. Flow diagram of the cooperative criteria assessment for game selection

After selecting the cooperative video game, it would be categorized under one of the four cooperative genres: puzzle platformer, asymmetric, simulation, and survival. The initial research plan targeted five video games per genre. Not all genres achieved five games due to technical requirements described next.

Technical requirements

After passing through the three cooperative criteria, researchers assessed whether there is enough footage for the selected game. Researchers decided on a minimum of seven and maximum 10 teams per video game. The footage selection criteria are covered in Table 10.

Table 10. Footage selection criteria

Criteria	**Requirements**
Verbal commentary	Clear verbal commentary with distinguishable player voices.
Game characters visibility	Shared or split screen where all player characters can be consistently observed.
Footage time	Follow genre specific rules to cover targeted number of levels or time.

Genre Footage Requirements

Since different games have different paces of mechanics, puzzles, and cooperative requirements, researchers followed genre specific rules to determine the footage time.

Puzzle platformer games are level-based games. Every level has several cooperative mechanics. Since puzzles are implemented in a 3D environment, players spend time exploring the environment and moving from one puzzle to another and therefore the footage does not always display cooperative requirements. Therefore, the channel must cover at least one complete level, and the time ranges between 20 to 40 minutes.

Asymmetric genre games are level-based games. Some games followed an asymmetric structure with 3D environments. Like puzzle platformer, players can explore the environment and move from one puzzle to another. Other games followed a puzzle-based approach, where players clear one puzzle at a time. The channel must cover at least one complete level or three complete puzzles and the time ranges between 20-40 minutes.

Simulation genre is fast paced level or puzzle based repetitive tasks. If puzzle based, every puzzle needs 3 to 5 minutes to complete, where players repeat the same game mechanic. The channel must cover at least 4 levels, and the time ranges between 10 to 20 minutes. If level based, the level could take 15-40 minutes to complete, and the channel must cover at least one level.

Survival genre is not level based. Therefore, the researchers chose a rule of 15-16 minutes. For consistency purposes, the 10 selected teams must be playing the same levels (unless it is not applicable, for example in Don’t starve together, levels can be randomly generated with every team). Table 11 summarizes the selected time/level for every genre.

Table 11. Selected time and levels for every genre

Genre	**Time/Level criteria**
Puzzle Platformer	• 20-40 minutes of gameplay • At least one full level
Asymmetric	• 20-40 minutes of gameplay • At least one full level (if level based) • At least 3 puzzles (if puzzle based)
Simulation	• 10-20 minutes of gameplay • At least 1 full level (if level based), at least three puzzles (if puzzle based).
Survival	• 15-16 minutes

Included games

Table 12 summarizes the genre descriptions and final lists of games that were included in the analysis.

Table 12. Genre summary and final list of video games included in the study

Genres	Games
Adventure Puzzle Platformer • Cooperative puzzles across the gameplay (Seif El-Nasr et al., 2010) • Variety of cooperative mechanics with mandatory interdependence • Puzzles are implemented in a 3D environment	It Takes Two (Hazelight Studios, 2021)
	Trine 4 (Forzenbyte, 2019)
	Portal 2 (Valve Corporation, 2011)
	Shift Happens (Klonk Games, 2015)
	BiPed (NexT Studios, 2020)
Survival: • The game is driven by collective survival and thus players are faced with a variety of risks (Sicart, 2015). • Mandatory and optional interdependencies of tasks so that players must play together to survive. • It supports crafting, collecting, and sharing.(Sicart, 2015)	Don't Starve Together (Klei Entertainment, 2016)
	The Survivalists (Team 17 & Mouldy Toof Studios, 2020)
	Grounded (Obsidian Entertainment, 2022)
	Atroneer (System Era Softworks, 2016)
Asymmetric: Players are presented with at least two of these asymmetries, and they are the core gameplay mechanic: interface, information, challenge, goals, responsibilities (Harris, 2019)	Keep Talking and Nobody Explodes (Steel Crate Games, 2015)
	We Were Here Forever (Total Mayhem Games & TMG Studios B.V., 2022)
	Tick Tock a tale for two (Other Tales Interactive, 2019)
	Operation Tango (Clever Plays, 2020)
Simulation: • Games that simulate tasks throughout the gameplay (Apperley, 2006). (Cooking simulation, train simulation, space shooter simulation etc.) • The game is reliant on performative engagement rather than tactical (such as puzzle solving) (Calleja, 2007)	Lovers in a dangerous spacetime (Astreoid Base, 2015)
	Overcooked 2 (Team 17 & Ghost Town Games, 2018)
	Unrailed (Indoor Atronaut, 2019)
	Catastronauts (Inertia Game Studios, 2018)
	KeyWe (Stonewheat & Sons, 2020)

Table 13 summarizes the number of teams, number of levels and average time coded for every video game. A total of 177 teams were coded. A total of 50 teams with an average time of 24 minutes were coded in puzzle platformer genre, 40 teams with an average time 26 minutes in asymmetric, 47 teams with an average time of 16 minutes in simulation and 40 teams with an average of 16 minutes in survival.

Table 13. Summary of teams, levels and average time coded for every video game

Video Game	Number Teams	Coded Levels	Average Time
Puzzle Platformer			
It Takes Two (Hazelight Studios,2021)	10	One Level	30 minutes
BiPed (NexT Studios, 2020)	10	Two Levels	28 minutes
Shift Happens (Klonk Games, 2015)	10	Four to Five Levels	27 minutes
Trine 4 (Forzenbyte, 2019)	10	One Level	21 minutes
Portal 2 (Valve Corporation, 2011)	10	Six Levels	15 minutes
Asymmetric			
Keep Talking and Nobody Explodes (Steel Crate Games, 2015)	10	Four to Six Levels	20 minutes
Operation Tango (Clever Plays, 2020)	10	One Level	23 minutes
Tick Tock a tale for two (Other Tales Interactive, 2019)	10	One Level	28 minutes
We Were Here Forever (Total Mayhem Games & TMG Studios B.V., 2022)	10	One Level	35 minutes
Simulation			
Overcooked 2 (Team 17 & Ghost Town Games, 2018)	10	Four to Five Levels	17 minutes
KeyWe (Stonewheat & Sons, 2020)	10	Four Levels	14 minutes
Catastronauts (Inertia Game Studios, 2018)	9	Five Levels	15 minutes
Lovers in a dangerous spacetime (Astreoid Base, 2015)	8	One Level	11 minutes
Unrailed (Indoor Atronaut, 2019)	10	Five to Seven Levels	23 minutes
Survival			
Don't Starve Together (Klei Entertainment, 2016)	10		16 minutes
The Survivalists (Team 17 & Mouldy Toof Studios, 2020)	10		15 minutes
Grounded (Obsidian Entertainment, 2022)	10		16 minutes
Atroneer (System Era Softworks, 2016)	10		16 minutes

Data Analysis Plan

The Data analysis plan aims to answer the three research questions. RQ1 aims to explore the consistency of video games in inducing teaming behaviors. To answer this question, homogeneity analysis was conducted within a genre to explore whether cooperative video games within the same cooperative genres, are consistent in inducing teaming profiles. RQ2 asks about the associations between cooperative features and teaming behaviors. The analysis is conducted through breaking down teaming competencies and cooperative features into percentages of associations. Finally, RQ3 targets the differences in behavioral markers' frequencies between upper and lower performers, to gain a better understanding of how teamwork behaviors frequency was affecting performance outcomes within every genre

RQ1: Consistency of Video Games in Inducing Teamwork Behaviors

Figure 6 presents the steps of analyzing the first research question.

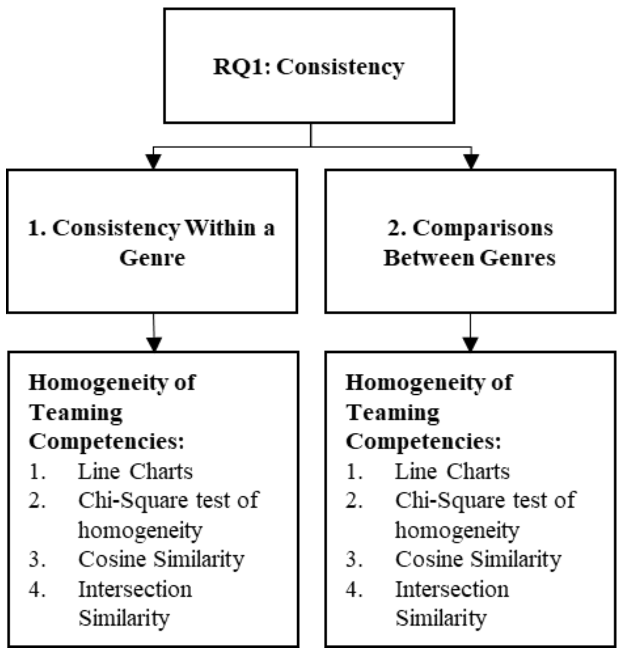


Figure 6. RQ1 analysis plan

Consistency within a Genre

To explore the consistency of the analyzed cooperative video games in inducing teaming behaviors, percentages and counts per minute were compared within genres. Since different video games were analyzed for different amounts of time, the results were normalized into percentages, by dividing every frequency of a teaming competency, by the total number of teamwork behaviors coded. The process of deriving percentages was conducted for every team in a video game, and the average was calculated with the standard error. The analysis involved chi-square homogeneity tests, cosine similarity, and intersection similarity for every genre.

Chi-square test of homogeneity

The chi-square test of homogeneity aims to answer whether the distribution of the teaming competencies is homogeneous between video games. First, an assessment of assumptions was conducted.

First, the chi-square analysis should be applied on a contingency table, where one categorical variable is assessed for homogeneity between two or more populations. In this study, the variable is the teaming competencies, and the populations constitute the video games (when comparing within genre), and the genres (when comparing between genres). Second, the data in the cells should be representing counts or frequencies. In this analysis, the data in the contingency table represent counts of every teaming competency. Third, the data should be mutually exclusive, hence one subject is not contributing to more than one level of the variable. In this study, the behavior was considered the subject of interest, and therefore a behavior should not be coded as two or more behavioral markers. Fourth, the data should be probabilistically independent. The most conservative approach is to extract each observation from a different subject to ensure independence (Wickens, 1989). This approach was not applied in this study. In this case, every team provided eight observations (every team was assessed for eight competencies), and within competencies repeated measures from the same subject were taken. However, in some cases it is applicable to assume independence of observations collected from one subject (Wickens, 1989). Hence, the underlying assumption to run this test assumes that the violation of independence is occasional, rather than a general violation that deems the test inadequate. The reasoning behind this assumption is based on the following.

First, even if one team contributed to eight competencies, the average of teaming competencies included in the contingency table is derived from independent teams. Therefore, the general distribution of competencies is assessed based on an average of independent teams.

Second, the most prominent dependencies were eliminated by organizing the data in higher dimension competencies, a method suggested by Wickens (1989). For example, it is assumed that it is more likely that behavioral markers within a competency are more dependent (e.g., the two behavioral markers SA-CR and SA-CM under situation assessment, or MA, DP, and TL under analysis and planning). By organizing the behavioral markers into higher dimensions of competencies, the most prominent dependencies were eliminated (dependencies of behavioral markers within every competency). Third, the observations within cells (repeated counts of same competencies), were assumed to be independent, since the underlying assumption of the study is that the behaviors are driven by the cooperative features (cooperative feature-behavior pairs). Since several similar measures are taken from the same subject, yet they are not assumed to be necessarily influenced by the subject, independence is assumed. Finally, chi-square is a commonly used method in corpus linguistics, where researchers assess the homogeneity of word distributions across different literature genres for example (Oakes, 2019; Gries, 2010; Baron 2009) in these cases, same words, and words across levels are not entirely independent since they can be generated from the same documents. In conclusion, using chi-square in this study argues for an occasional violation of interdependence, rather than a general violation that deems the test unsuitable for this use case. The chi-square test is followed by similarity measures to further explore homogeneity.

The chi-square test reports the chi-square test statistic, p-value, and Cramer's v. Cramer's v is an effect size generated with the chi-square test statistic. Cramer's v's thresholds are affected by the degree of freedom. To determine which thresholds to use, we subtract 1 from number of columns and number of rows, and take the minimum (Pallant, 2011) (in this case number of video games – 1 or number of genres – 1). This approach follows the rules of thumbs provided for Cohen’s w (with 0.1 small, 0.3 medium and 0.5 large) (Cohen, 1969). Cramer’s v is equivalent to Cohen’s w multiplied by square root of the minimum obtained in the described step (e.g., a Cohen’s w of 0.1 is equal to Cramer’s v of 0.06 when c-1 is 3, and 0.05 when c-1 is 4). Table 14 summarizes the thresholds.

Table 14 Cramer's v thresholds

Cramer’s v thresholds	**c-1=4**	**c-1=3**
Small	0.05	0.06
Medium	0.15	0.17
Large	0.25	0.29

Similarity measures

Cosine Similarity

Similar measures were calculated for both percentages and counts per minute to further assess how similar video games within genres are (Cha, 2008). The applied similarity measures can be applied to nominal or categorical histogram, which is applicable in this study. In a nominal type histogram, the ordering of the levels does not affect the similarity measures (Cha & Srihari, 2002). For example, the similarity measure between two video games will not change if the order of the competencies is changed, if they both follow the same order.

Cosine similarity measures the angular metric between two vectors, and it is calculated by dividing the inner product of two vectors by the product of their magnitudes (Cha, 2008). Cosine similarity determines whether two vectors point in the same direction and is commonly used in document similarity (Han et al., 2012). In document similarity, a document is turned into a vector, by counting the occurrences of a term in the document (Han et al., 2012). The cosine similarity would be used to measure how similar the two documents are, by comparing the two vectors. In this study, the video game footage would be considered the unit of comparison. The video is turned into a vector constituting of counts of occurrences of teaming behaviors. Therefore, the cosine similarity would assess how similar the video game's content is in terms of teaming behaviors. A cosine value of zero means the two vectors are perpendicular and have no match. The closer the cosine is to one, the smaller the angle and the higher the match between the vectors (Han et al., 2012). In this study, the cosine similarity value was used to assess how similar the distribution of teaming behaviors is within genres, and between genres.

Intersection Similarity

Intersection similarity measures the overlap between the two vectors, by summing the minimum of every teaming competency percentage or count (Cha, 2008). In this study, the intersection was divided by the union, to obtain a proportion of how much the overlap constitutes of the union of both vectors. While the cosine similarity is equal to one when two vectors are pointing in the same direction, intersection similarity is used to measure the overlap between the two vectors, and therefore would be equal to 1 when two vectors completely overlap. Therefore, the higher the overlap, the higher the intersection measure. While cosine similarity is more robust to magnitude differences, intersection similarity can better reflect that.

Comparisons between Genres

This section presents results that aim to compare the teaming competencies' profiles between genres. The same steps followed for consistency assessment within genres were followed (chi-square, cosine, intersection), however video games within a genre were now considered the same population, representing one genre.

RQ2: Associations Between Cooperative Features and Teaming Behaviors

Figure 7 summarizes the analysis plan for RQ2.

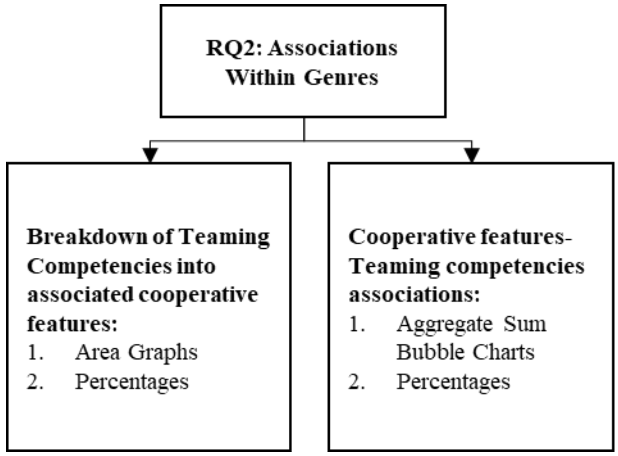


Figure 7. RQ2 analysis plan

Breakdown of teaming competencies into associated cooperative features

To gain a better understanding what cooperative features were inducing the teaming competencies within every genre, area graphs were plotted, breaking down the teaming competencies into their cooperative features' compositions. This analysis aims to provide an understanding of the underlying cooperative features that contributed to the teaming profile in every genre. The area graphs present categorical breakdowns of every teaming competency to represent what cooperative features were involved in inducing it. An aggregate sum of cooperative feature-behavioral marker pairs was calculated within every genre, summing across all video games and teams within a genre.

Cooperative features-teaming competencies associations

To explore what behavioral markers are most frequently associated with cooperative features, bubble charts representing the counts of cooperative feature-behavioral marker pairs are presented as a visual representation of the load of associations within every genre. This visualization explores why certain behavioral markers are dominant in certain genres, and what cooperative features were associated with them. Additionally, top three competencies associated with every cooperative feature will be summarized, to highlight what competencies are most frequently associated with the feature, therefore guiding teamwork testbed design recommendations in following chapters.

RQ3: Team behaviors' comparisons between upper and lower performers

Figure 8 summarizes the analysis plan for RQ3.

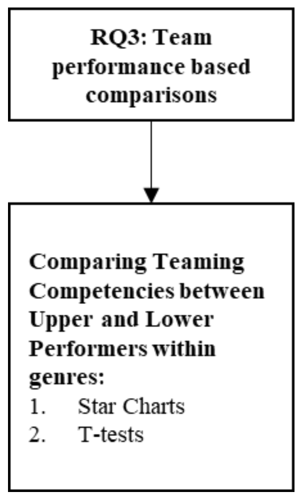


Figure 8. RQ3 analysis plan

The next analysis section aims to explore the behavioral differences between higher and lower performers. Teamwork behaviors are considered mediators for performance outcomes (Ilgen et al., 2005). This research question aims to explore if there are differences in the frequency of teamwork behaviors between upper and lower performers. In video games, Mayer (2018) suggested that empirically proving the relationship between teamwork activities and gaming performance outcomes contributes to proving the internal validity of video games as suitable teamwork assessment environments. This implies that teamwork in these gaming environments matters, and is actively influencing the performance outcomes, rather than the outcome solely depending on individual players skills and expertise for example. Hence, a valid teamwork cooperative game would be designed in a way that supports teamwork behaviors, which are necessary for successful performance outcomes.

Behavioral differences within a genre

For every video game within one genre, teams were split into upper and lower half based on their performance outcomes. The frequency of behavioral markers per minute of gameplay was calculated for every team. Afterwards, upper and lower halves from every video game within the genre were compiled, to compare the two performance categories. Every teaming competency was compared through a t-test to assess whether there exists a statistically significant difference between upper and lower half performers. To perform t-tests, assumptions were first reviewed. First, frequency was considered continuous, since it represented the count of behaviors divided by time, which is a continuous variable. Second, Shapiro-wilk test of normality was run for upper and lower teamwork frequencies for every competency. When Shapiro-wilk was statistically significant, the Q-Q plots were reviewed. If no major deviations from the diagonal were observed, the t-test was run. No transformations were made to the data to account for normality assumption violations.

Teams within puzzle platformer and asymmetric genres were separated based on time to finish the level and total failures. Two teams were excluded from the team performance analysis in asymmetric genre due to streamers skipping time from the gameplay. Therefore, comparing their count of behaviors per minute with other teams, based on their performance was deemed not reflective of how the team engage in team behaviors. Teams in simulation genre were separated based on scores provided by the games. Unrailed was not included in the performance comparisons since there was no clear way to separate upper and lower performers. Levels in unrailed are randomly generated, hence overall teams did not play same levels, and the performance outcomes can vary depending on teams' approaches. Finally, a performance comparison was not conducted for survival genre. Since survival genre is not level oriented (e.g., clearing puzzles, winning levels), there was no clear threshold on how to separate the teams.

CHAPTER 4. CONSISTENCY OF COOPERATIVE VIDEO GAMES IN INDUCING TEAMWORK BEHAVIORS (RQ1)

This chapter presents the results and discussion of the analysis to assess the consistency of cooperative games in inducing teamwork behaviors (RQ1). This analysis aims to explore the consistency of the distribution of teamwork behaviors induced within the four analyzed genres: puzzle platformer, asymmetric, simulation and survival. By investigating the patterns of teamwork competencies within genres, the analysis aims to provide a predictive approach of how teamwork competencies are distributed within genres, and how designers can use cooperative genres to target teamwork competencies. Furthermore, through investigating the differences between genres, the analysis aims to explore how cooperative genres affect the distributions of teamwork behaviors, and therefore, how to use different cooperative genres in testbed design to emphasize different teamwork distributions.

The results are followed with an RQ1 discussion, reflecting on the findings. The discussion elaborates on the consistency within genres and the differences between genres and how they can be used as testbed design approaches to target teamwork profiles.

Puzzle Platformer

Consistency of Competencies

Figure 9 illustrates the average percentages and standard error of teaming competencies for the puzzle platformer genre, comparing five games: BiPed (BP), It Takes Two (ITT), Portal 2 (P2), Shift Happens (SH), and Trine 4 (T4). Percentages of competencies were calculated for every team and averaged within every video game.

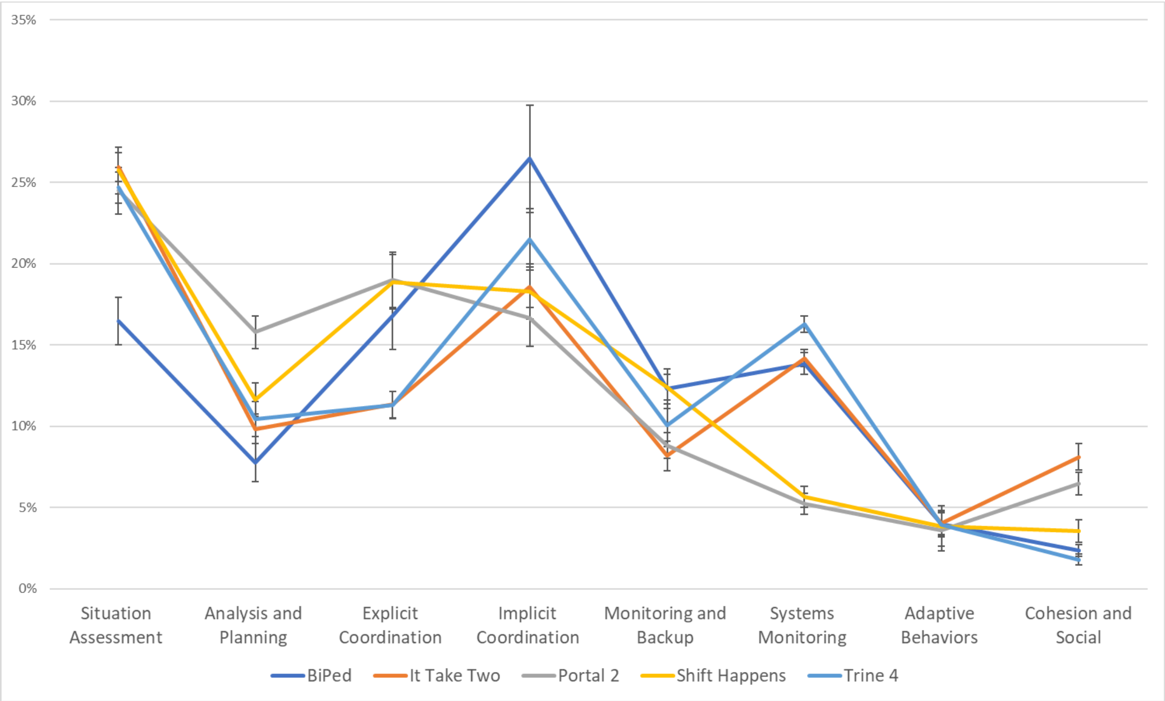


Figure 9. Percentages of teaming competencies within puzzle platformer genre

Table 15 shows the average and standard error of the five puzzle platformer video games.

Table 15. Average and standard error of the percentages of teaming competencies within puzzle platformer genre.

Competency	BP	ITT	P2	SH	T4	AVG
Situation Assessment	16.5 (1.5)	25.9 (0.9)	24.5 (1.4)	25.7 (0.9)	24.7 (0.9)	23.5 (2.0)
Analysis/Planning	7.7 (1.2)	9.8 (0.9)	15.8 (1.0)	11.6 (1.1)	10.4 (0.9)	11.1 (1.5)
Explicit Coordination	16.8 (2.1)	11.3 (0.8)	19.0 (1.7)	18.9 (0.9)	11.3 (1.2)	15.5 (1.9)
Implicit Coordination	26.5 (3.3)	18.6 (1.2)	16.6 (1.7)	18.3 (1.9)	21.5 (1.5)	20.3 (1.9)
Monitoring/Backup	12.3 (1.2)	8.2 (0.9)	8.8 (0.8)	12.4 (1.3)	10.1 (1.1)	10.4 (1.0)
Systems Monitoring	13.9 (0.7)	14.2 (0.5)	5.2 (0.6)	5.6 (0.5)	16.3 (0.7)	11.0 (2.6)
Adaptive Behaviors	4.0 (0.8)	4.0 (0.7)	3.6 (1.2)	3.9 (0.7)	3.9 (0.6)	3.9 (0.1)
Cohesion/Social	2.4 (0.4)	8.1 (0.8)	6.5 (0.7)	3.5 (0.3)	1.8 (0.3)	4.5 (1.4)

Chi Square test of Homogeneity

The Pearson Chi Square statistic was $X^2(28) = 41.151$ (p=.022), and Cramer's v = 0.123, which is considered small effect size. Table 16 provides the letter report for the Bonferroni adjusted post-hoc analysis. Systems monitoring was significantly different between the five video games, with Portal 2 having the lowest percentage (5.7%) and Trine 4 having the highest percentage of 15.7%.

Table 16. Post-hoc pairwise comparisons within puzzle platformer genre

		Puzzle Platformer Video Game				
		BP	ITT	P2	SH	T4
Teaming Competency	SA	A	A	A	A	A
	AP	A	A	A	A	A
	EC	A	A	A	A	A
	IC	A	A	A	A	A
	MB	A	A	A	A	A
	SM*	AB	AB	B	AB	A
	AB	A	A	A	A	A
	CS	A	A	A	A	A

Cosine Similarity

Table 17 shows the cosine similarity values between every pair of video games in Puzzle Platformer genre, with a total average of 0.94 (0.01) which is considered close to 1, the maximum value of cosine similarity.

Table 17. Cosine similarity measures between puzzle platformer video games (Percentages)

Cosine Similarity	**ITT**	**P2**	**SH**	**T4**	**AVG (SE)**
BP	0.93	0.90	0.93	0.96	
ITT		0.94	0.95	0.98	
P2			0.99	0.92	
SH				0.94	
					0.94 (0.01)

Intersection Similarity

Table 18 shows the intersection over union similarity values between every pair of video games. On average, two puzzle games overlap 75% (SE = 2%) of their total union.

Table 18. Intersection over union measures between puzzle platformer video games (Percentages)

Intersection over Union	ITT	P2	SH	T4	AVG (SE)
BP	0.70	0.63	0.72	0.76	
ITT		0.75	0.76	0.86	
P2			0.87	0.70	
SH				0.76	
					0.75 (0.02)

Team Behaviors Frequency

Figure 10 shows the line charts of the counts of competencies per minute within puzzle platformer genre.

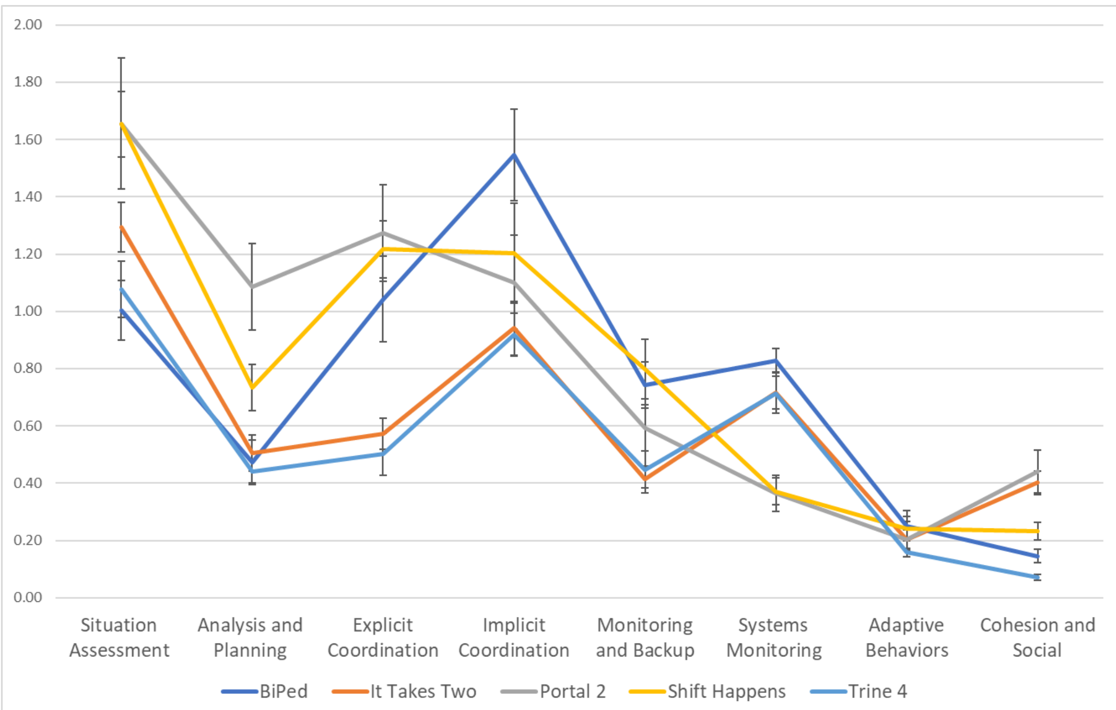


Figure 10. Frequencies of teaming competencies within puzzle platformer genre

Table 21 summarizes the average and standard error of counts/minute of teaming competencies for puzzle platformer video games.

Table 19 Frequency of teaming competencies within puzzle platformer genre

Competency	BP	ITT	P2	SH	T4	AVG
Situation Assessment	1.00 (0.10)	1.29 (0.08)	1.65 (0.22)	1.65 (0.11)	1.08 (0.09)	1.34 (0.14)
Analysis Planning	0.47 (0.08)	0.51 (0.06)	1.09 (0.15)	0.73 (0.08)	0.44 (0.04)	0.65 (0.12)
Explicit Coordination	1.04 (0.15)	0.57 (0.05)	1.27 (0.17)	1.22 (0.10)	0.50 (0.07)	0.92 (0.16)
Implicit Coordination	1.55 (0.16)	0.94 (0.09)	1.1 (0.16)	1.2 (0.17)	0.92 (0.07)	1.14 (0.11)
Monitoring and Backup	0.74 (0.08)	0.41 (0.04)	0.59 (0.08)	0.80 (0.10)	0.45 (0.06)	0.6 (0.08)
Systems Monitoring	0.83 (0.04)	0.72 (0.06)	0.36 (0.06)	0.37 (0.05)	0.71 (0.07)	0.6 (0.1)
Adaptive Behaviors	0.25 (0.05)	0.20 (0.04)	0.20 (0.06)	0.24 (0.04)	0.16 (0.02)	0.21 (0.02)
Cohesion and Social	0.15 (0.02)	0.40 (0.04)	0.44 (0.07)	0.23 (0.03)	0.07 (0.01)	0.26 (0.07)

Cosine Similarity

Table 20 summarizes the cosine similarity values between every pair of video games in Puzzle Platformer genre, comparing frequencies, with a total average of 0.94 (0.02) which is considered close to one, the maximum value of cosine similarity.

Table 20. Cosine similarity measures between puzzle platformer video games (Frequencies)

Cosine Similarity	ITT	P2	SH	T4	AVG (SE)
BP	0.93	0.90	0.93	0.95	
ITT		0.94	0.95	0.98	
P2			0.98	0.92	
SH				0.94	
					0.95 (0.01)

Intersection Similarity

Table 21 shows the Intersection over Union values between every pair of video games in Puzzle Platformer genre, comparing frequencies, with an average of 0.68 (0.03).

Table 21. Intersection over union measures between puzzle platformer video games (Frequencies)

Intersection over Union	ITT	P2	SH	T4	AVG (SE)
BP	0.68	0.63	0.72	0.70	
ITT		0.66	0.65	0.85	
P2			0.86	0.58	
SH				0.43	
					0.68 (0.04)

Table 22 summarizes the similarity measures within puzzle platformer, when compared for percentages of teamwork competencies, and for counts per minute.

Table 22. Similarity measures' averages within puzzle platformer genre

Similarity Measure	Percentages	Counts per minute
Cosine Similarity	0.94 (0.01)	0.95 (0.01)
Intersection Similarity	0.75 (0.02)	0.68 (0.04)

Asymmetric

Consistency of Competencies

Figure 11 illustrates the average percentages and standard error of teaming competencies for the asymmetric genre: Keep Talking and Nobody Explodes (KT), We Were Here Forever (WWHF), Tick Tock a Tale for Two (TT) and Operation Tango (OT).

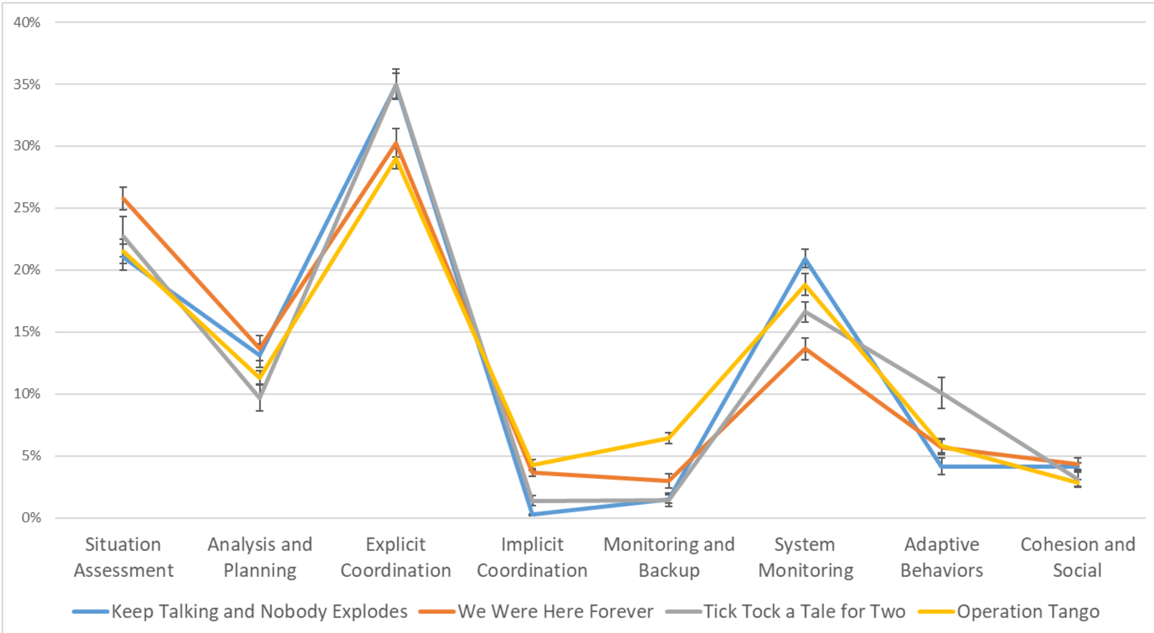


Figure 11. Percentages of teaming competencies with asymmetric genre

Table 23 shows average (AVG) and standard error (SE) of the four Asymmetric games.

Table 23. Average percentages and standard error of teaming competencies within asymmetric genre

Competency	**KT**	**WWHF**	**TT**	**OT**	**AVG**
Situation Assessment	21.0 (1.1)	25.8 (0.9)	22.7 (1.6)	21.5 (1.0)	22.8 (1.1)
Analysis and Planning	13.1 (0.9)	13.7 (1.0)	9.7 (1.1)	11.3 (0.6)	11.9 (0.9)
Explicit Coordination	34.9 (1.0)	30.2 (1.1)	35.0 (1.2)	29.0 (0.8)	32.3 (1.6)
Implicit Coordination	0.2 (0.1)	3.6 (0.3)	1.4 (0.4)	4.3 (0.5)	2.4 (0.9)
Monitoring and Backup	1.5 (0.3)	3.0 (0.6)	1.5 (0.5)	6.4 (0.4)	3.1 (1.2)
Systems Monitoring	20.9 (0.7)	13.7 (0.9)	16.6 (0.8)	18.8 (0.9)	17.5 (1.6)
Adaptive Behaviors	4.2 (0.7)	5.7 (0.6)	10.1 (1.2)	5.8 (0.6)	6.4 (1.3)
Cohesion and Social	4.2 (0.3)	4.3 (0.5)	3.1 (0.6)	2.9 (0.3)	3.6 (0.4)

Chi Square test of Homogeneity

A contingency table analysis was conducted to assess the homogeneity of the distribution of the teaming behaviors within Asymmetric video games genre. The Pearson Chi Square statistic $X^2(21) = 30.863$ ($p = .076$), considered not statistically significant. Table 24 provides the letter report for the post-hoc analysis.

Table 24. Post-hoc pairwise comparisons within asymmetric genre

		Asymmetric Video Games			
		KT	OT	TT	WWH
Teaming Competency	SA	A	A	A	A
	AP	A	A	A	A
	EC	A	A	A	A
	IC*	A	B	AB	AB
	MB	A	A	A	A
	SM	A	A	A	A
	AB	A	A	A	A
	CS	A	A	A	A

Cosine similarity

Table 25 shows the cosine similarity values between every pair of video games in asymmetric genre, with a total average of 0.982 (0.001) which is considered close to 1

Table 25. Cosine Similarity measures between asymmetric video games (Percentages)

Cosine Similarity	**WWHF**	**TT**	**OT**	**AVG (SE)**
KT	0.976	0.985	0.984	
WWHF		0.981	0.984	
TT			0.980	
				0.982 (0.001)

Intersection similarity

Table 26 shows the intersection similarity values with an average of 0.8 (0.01)

Table 26. Intersection over union measures between asymmetric video games (Percentages)

Intersection Similarity	**WWHF**	**TT**	**OT**	**AVG (SE)**
KT	0.78	0.84	0.80	
WWHF		0.78	0.82	
TT			0.78	
				0.8 (0.01)

Team Behaviors Frequency

Figure 12 shows the line charts of the average counts of competencies per minute and standard error within asymmetric genre.

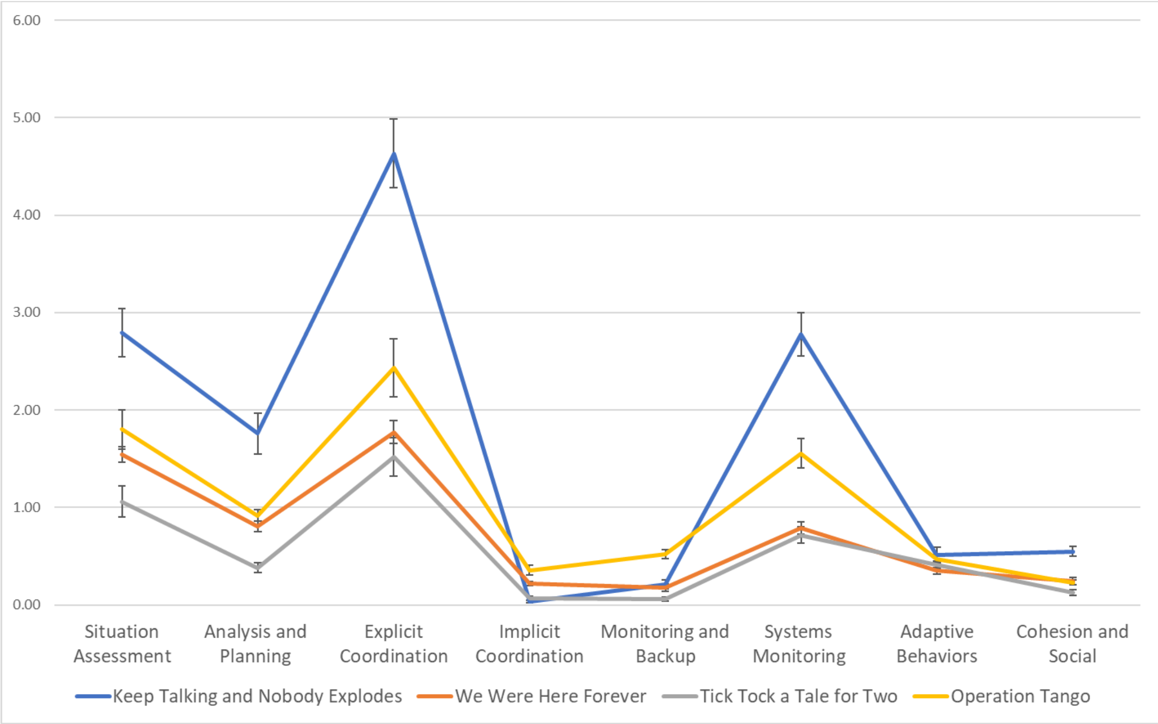


Figure 12. Frequencies of teaming competencies within asymmetric genre

Table 27 summarizes the average and standard error of counts/minute.

Table 27. Frequency of teaming competencies within asymmetric genre

Competency	**KT**	**WWHF**	**TT**	**OT**	**AVG**
Situation Assessment	2.79 (0.25)	1.54 (0.08)	1.06 (0.16)	1.8 (0.2)	1.8 (0.73)
Analysis/Planning	1.76 (0.21)	0.81 (0.06)	0.38 (0.05)	0.92 (0.06)	0.97 (0.58)
Explicit Coordination	4.63 (0.35)	1.77 (0.12)	1.52 (0.2)	2.43 (0.29)	2.59 (1.41)
Implicit Coordination	0.03 (0.01)	0.22 (0.02)	0.07 (0.02)	0.36 (0.05)	0.17 (0.15)
Monitoring/ Backup	0.21 (0.05)	0.18 (.04)	0.06 (0.02)	0.52 (0.04)	0.24 (0.2)
Systems Monitoring	2.78 (0.22)	0.79 (0.06)	0.72 (0.09)	1.55 (0.15)	1.46 (0.96)
Adaptive Behaviors	0.51 (0.08)	0.35 (0.03)	0.41 (0.04)	0.47 (0.05)	0.44 (0.07)
Cohesion/Social	0.55 (0.05)	0.24 (0.04)	0.13 (0.03)	0.23 (0.03)	0.29 (0.18)

Cosine Similarity

Table 28 summarizes cosine similarity values for counts/minute within asymmetric.

Table 28. Cosine similarity measures between asymmetric video games (Frequencies)

Cosine Similarity	WWHF	TT	OT	AVG (SE)
KT	0.973	0.981	0.984	
WWHF		0.982	0.983	
TT			0.981	
				0.981 (0.002)

Intersection Similaritycounts/minute within asymmetric.

Table 29 summarizes the intersection similarity for counts/minute within asymmetric.

Table 29. Intersection over union measures between asymmetric video games (Frequencies)

Intersection over Union	WWHF	TT	OT	AVG (SE)
KT	0.42	0.32	0.55	
WWHF		0.72	0.71	
TT			0.52	
				0.54 (0.06)

Table 30 summarizes the cosine and intersection similarity for percentages and counts.

Table 30. Similarity measures' averages within asymmetric genre

Similarity/Distance Measure	**Percentages**	**Count per minute**
Cosine Similarity	0.982 (0.001)	0.981 (0.002)
Intersection Similarity	0.804 (0.009)	0.54 (0.06)

Simulation

Consistency of Competencies

Figure 13 illustrates the average percentages of teaming competencies for the simulation genre, comparing five games: KeyWe (KW), Overcooked 2 (OC2), Unrailed (UR), Catastronauts (CN) and Lovers in a Dangerous Spacetime (LDST).

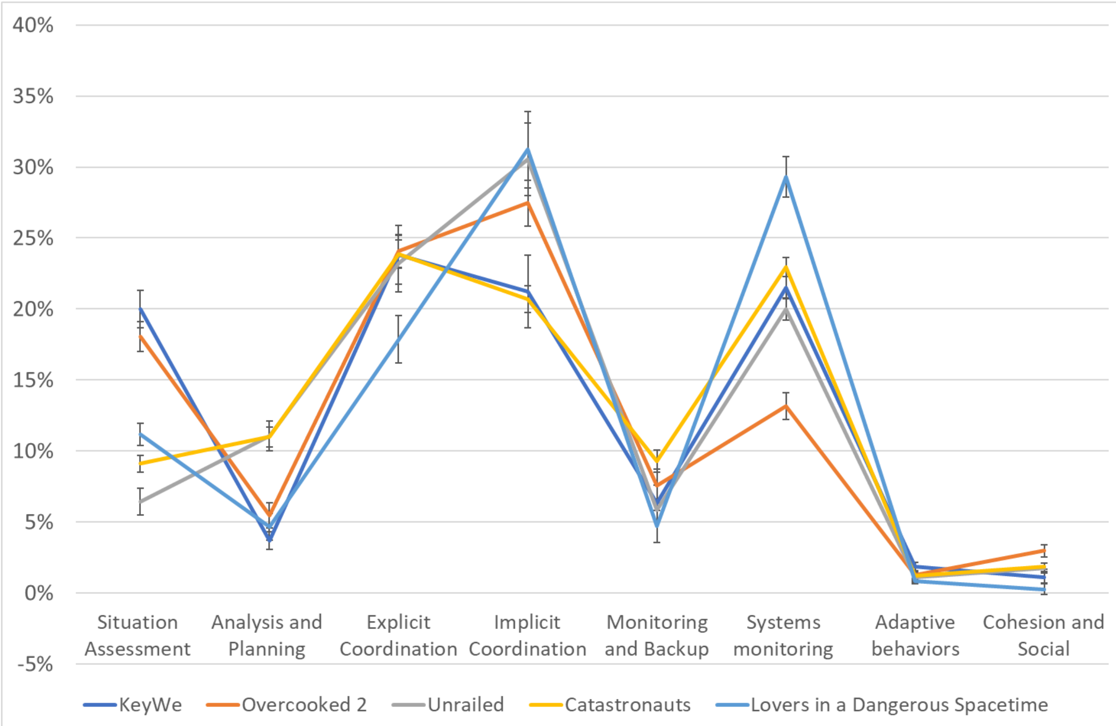


Figure 13. Percentages of teaming competencies within simulation genre

Table 31 shows average (AVG) and standard error (SE) of the five simulation video games. The results are reported as AVG (SE).

Table 31. Average percentages and standard error of teaming competencies within puzzle simulation genre

Competency	KW	OC2	UR	CN	LDST	AVG
Situation Assessment	20.0 (1.3)	18.0 (1.0)	6.4 (0.9)	9.1 (0.6)	11.2 (0.8)	13.0 (2.6)
Analysis/Planning	3.7 (0.6)	5.5 (0.9)	11.1 (1.1)	11.0 (0.7)	4.7 (0.9)	7.2 (1.6)
Explicit Coordination	23.8 (2.1)	24.1 (1.2)	23.2 (2.0)	23.9 (1.0)	17.9 (1.7)	22.6 (1.2)
Implicit Coordination	21.1 (2.5)	27.4 (1.6)	30.5 (2.6)	20.7 (0.9)	31.2 (2.7)	26.2 (2.2)
Monitoring/Backup	6.3 (1.3)	7.6 (1.2)	5.9 (0.8)	9.3 (0.8)	4.7 (1.1)	6.7 (0.8)
Systems Monitoring	21.5 (0.8)	13.1 (0.9)	20.0 (0.8)	22.9 (0.7)	29.3 (1.4)	21.4 (2.6)
Adaptive Behaviors	1.8 (0.4)	1.3 (0.4)	1.1 (0.3)	1.2 (0.2)	0.8 (0.4)	1.3 (0.2)
Cohesion/Social	1.1 (0.3)	3.0 (0.5)	1.8 (0.3)	1.9 (0.3)	0.3 (0.2)	1.6 (0.4)

Chi-square test of Homogeneity

A contingency table analysis was conducted to assess the homogeneity of the distribution of the teaming behaviors within Simulation genre. Chi Square statistic $X^2(28) = 66.978$ *($p<.001$)*, and Cramer’s v of 0.131 which is considered a small effect size. Table 32 summarizes the results of Bonferroni post hoc test comparing proportions between genres for every competency. Situation assessment and systems monitoring were significantly different within the simulation genre.

Table 32. Post-hoc pairwise comparisons within simulation genre

		Simulation Video Game				
		CTN	KW	LDST	OC2	UR
Teaming Competency	SA*	AB	C	ABC	BC	A
	AP	A	A	A	A	A
	EC	A	A	A	A	A
	IC	A	A	A	A	A
	MB	A	A	A	A	A
	SM*	AB	AB	B	A	AB
	AB	A	A	A	A	A
	CS	A	A	A	A	A

Cosine Similarity

Table 33 shows the cosine similarity values between every pair of video games, with a total average of 0.94 (0.01) which is considered close to 1, the maximum value of cosine similarity.

Table 33. Cosine similarity measures between simulation video games (Percentages)

Cosine Similarity	**OC2**	**UR**	**CN**	**LDST**	**AVG (SE)**
KW	0.97	0.92	0.95	0.94	
OC2		0.94	0.93	0.92	
UR			0.97	0.96	
CN				0.95	
					0.94 (0.01)

Intersection Similarity

Table 34 shows the intersection similarity values between every pair of video games, with an average of 0.73 (0.02)

Table 34. Intersection over union measures between simulation video games (Percentages)

Intersection over Union	**OC2**	**UR**	**CN**	**LDST**	**AVG (SE)**
KW	0.80	0.70	0.78	0.68	
OC2		0.73	0.71	0.67	
UR			0.82	0.74	
CN				0.68	
					0.73 (0.02)

Team Behaviors Frequency

Figure 14 illustrates the average frequencies behaviors and standard error for every teaming competency for the simulation genre: KeyWe (KW), Overcooked 2 (OC2), Unrailed (UR), Catastronauts (CN) and Lovers in a Dangerous Spacetime (LDST).

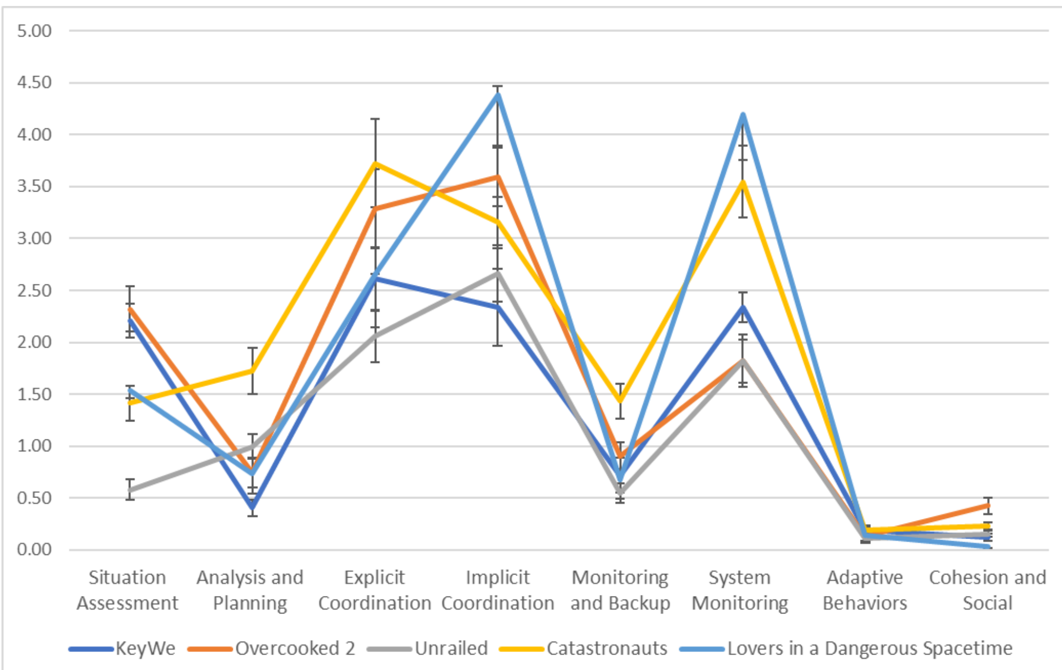


Figure 14 Frequencies of teaming competencies within simulation genre

Table 35 summarizes the frequency of teamwork behaviors for every teaming competency within simulation genre.

Table 35. Frequency of teaming competencies within simulation genre

Competency	KW	OC2	UR	CTNT	LDST	AVG
Situation Assessment	2.21 (0.16)	2.32 (0.22)	0.50 (0.07)	1.41 (0.16)	1.54 (0.08)	1.6 (0.33)
Analysis Planning	0.41 (0.07)	0.75 (0.14)	0.88 (0.12)	1.72 (0.23)	0.73 (0.19)	0.90 (0.22)
Explicit Coordination	2.61 (0.3)	3.28 (0.25)	1.84 (0.21)	3.73 (0.42)	2.65 (0.51)	2.82 (0.32)
Implicit Coordination	2.34 (0.37)	3.59 (0.28)	2.35 (0.18)	3.15 (0.25)	4.39 (0.49)	3.16 (0.39)
Monitoring and Backup	0.72 (0.17)	0.91 (0.13)	0.46 (0.06)	1.43 (0.16)	0.68 (0.18)	0.84 (0.16)
Systems Monitoring	2.33 (0.14)	1.82 (0.25)	1.52 (0.14)	3.55 (0.35)	4.2 (0.43)	2.70 (0.50)
Adaptive Behaviors	0.19 (0.03)	0.13 (0.04)	0.09 (0.03)	0.19 (0.04)	0.14 (0.07)	0.15 (0.02)
Cohesion and Social	0.12 (0.03)	0.42 (0.08)	0.14 (0.02)	0.23 (0.04)	0.04 (0.02)	0.19 (0.07)

Cosine Similarity

Table 36 summarizes the cosine similarity measures between the five simulation video games, with an average of 0.95 (0.01).

Table 36. Cosine similarity measures between simulation video games (frequencies)

Cosine Similarity	OC2	UR	CN	LDST	AVG (SE)
KW	0.97	0.92	0.95	0.94	
OC2		0.95	0.94	0.92	
UR			0.97	0.96	
CN				0.95	
					0.95 (0.01)

Intersection Similarity

Table 37 presents the intersection similarity measures between the five games within simulation genre, with an average of 0.63 (0.03).

Table 37. Intersection over union measures between simulation video games (frequencies)

Intersection over Union	**OC2**	**UR**	**CN**	**LDST**	**AVG (SE)**
KW	0.75	0.64	0.62	0.66	
OC2		0.58	0.69	0.68	
UR			0.51	0.49	
CN				0.71	
					0.63 (0.03)

Table 38 summarizes the average of all similarity measures within simulation.

Table 38. Similarity measures' averages within simulation genre

Similarity/Distance Measure	**Percentages**	**Frequency**
Cosine Similarity	0.94 (0.01)	0.95 (0.01)
Intersection over Union	0.73 (0.02)	0.63 (0.03)

Survival

Consistency of Competencies

Figure 15 shows the average percentages and standard error of teaming competencies: The Survivalists, Don't Starve Together, Astroneer, and Grounded.

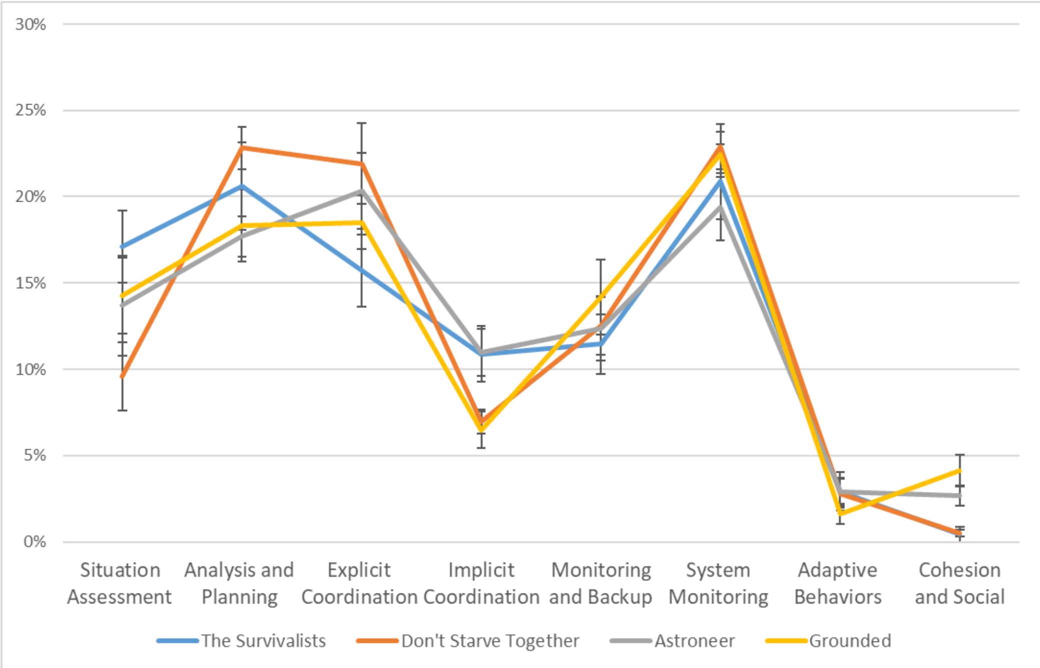


Figure 15. Percentages of teaming competencies within survival genre

Table 39 shows average (AVG) and standard error (SE) of the four survival genres: The Survivalists (SUR), Don't Starve Together (DST), Astroneer (ASTN) and Grounded (GRND). The results are reported as AVG (SE).

Table 39. Average percentages and standard error of teaming competencies within puzzle survival genre

Competency	**SUR**	**DST**	**ASTN**	**GRND**	**AVG**
Situation Assessment	17.1 (2.1)	9.6 (2.0)	13.7 (2.9)	14.3 (2.2)	13.7 (1.6)
Analysis/Planning	20.6 (2.5)	22.8 (1.2)	17.7 (1.2)	18.3 (2.1)	19.9 (1.2)
Explicit Coordination	15.7 (2.1)	21.9 (2.3)	20.3 (2.2)	18.5 (1.6)	19.1 (1.3)
Implicit Coordination	10.9 (1.6)	7.0 (0.7)	11.0 (1.3)	6.5 (1.0)	8.8 (1.2)
Monitoring/Backup	11.5 (1.7)	12.5 (1.7)	12.4 (1.9)	14.2 (2.2)	12.6 (0.6)
Systems Monitoring	20.9 (2.2)	22.9 (1.3)	19.4 (1.9)	22.4 (1.3)	21.4 (0.8)
Adaptive Behaviors	2.9 (1.1)	2.8 (0.8)	2.9 (0.8)	1.6 (0.6)	2.6 (0.3)
Cohesion/Social	0.4 (0.4)	0.5 (0.2)	2.7 (0.6)	4.2 (0.9)	1.9 (0.9)

Chi Square Homogeneity

A contingency table analysis was conducted to assess the homogeneity of the distribution of the teaming behaviors within Simulation genre. The Pearson Chi Square statistic $X^2(21) = 9.925$ (p=.970) and a Cramer's V of 0.102, which is considered small effect size. Table 40 illustrates the no significant differences with matching letters for all competencies.

Table 40. Post-hoc pairwise comparisons within survival genre

		Survival Video Games			
Teaming Competency		ASTN	DST	GRND	SUR
	SA	A	A	A	A
	AP	A	A	A	A
	EC	A	A	A	A
	IC	A	A	A	A
	MB	A	A	A	A
	SM	A	A	A	A
	AB	A	A	A	A
	CS	A	A	A	A

Cosine Similarity

Table 41 shows the cosine similarity values between every pair of video games in survival genre, with a total average of

Table 41. Cosine Similarity measures between survival video games (Percentages)

Cosine Similarity	DST	ASTN	GRND	AVG (SE)
SUR	0.97	0.99	0.98	
DST		0.98	0.98	
ANT			0.99	
				0.98 (0.003)

Intersection Similarity

Table 42 shows the intersection similarity values between every pair of video games in survival genre, with an average of 0.82 (.01)

Table 42. Intersection over union measures between survival video games (Percentages)

Intersection Similarity	DST	ASTN	GRND	AVG (SE)
SUR	0.79	0.86	0.81	
DST		0.81	0.82	
ANT			0.86	
				0.82 (0.01)

Team Behaviors Frequency

Figure 16 illustrates the frequencies of teamwork behaviors for every teaming competency in the survival genre: The Survivalists (SUR), Don’t Starve Together (DST), Astroneer (ASTN) and Grounded (GRND).

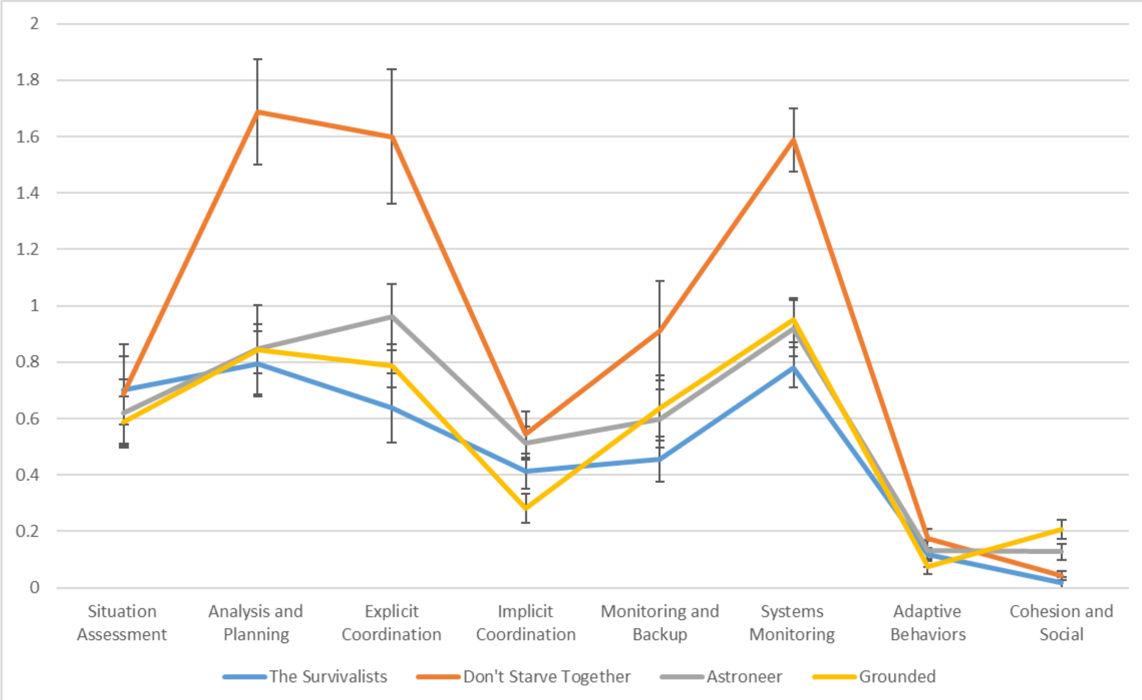


Figure 16. Frequencies of teaming competencies within survival genre

Table 43 summarizes the average and standard errors for frequencies of teamwork competencies.

Table 43. Frequency of teaming competencies within survival genre

Competency	**SUR**	**DST**	**ASTN**	**GRND**	**AVG**
Situation Assessment	0.70 (0.12)	0.69 (0.18)	0.62 (0.12)	0.59 (0.09)	0.65 (0.03)
Analysis/Planning	0.79 (0.12)	1.69 (0.19)	0.85 (0.09)	0.84 (0.16)	1.04 (0.22)
Explicit Coordination	0.64 (0.12)	1.6 (0.24)	0.96 (0.12)	0.79 (0.08)	1.00 (0.21)
Implicit Coordination	0.41 (0.06)	0.54 (0.08)	0.51 (0.06)	0.28 (0.05)	0.44 (0.06)
Monitoring/Backup	0.46 (0.08)	0.91(0.18)	0.6 (0.1)	0.64 (0.11)	0.65 (0.1)
System Monitoring	0.78 (0.07)	1.59 (0.11)	0.92 (0.1)	0.95 (0.08)	1.06 (0.18)
Adaptive Behaviors	0.12 (0.05)	0.18 (0.03)	0.13 (0.04)	0.08 (0.03)	0.13 (0.02)
Cohesion/Social	0.02 (0.02)	0.04 (0.02)	0.13 (0.03)	0.21 (0.03)	0.1 (0.04)

Cosine Similarity

Table 44 presents the cosine similarity measures between every pair of the four survival video games with an average of 0.98 (0.003).

Table 44. Cosine similarity values between survival video games (frequencies)

Cosine Similarity	DST	ASTN	GRND	AVG (SE)
SUR	0.98	0.98	0.99	
DST		0.96	0.98	
ANT			0.98	
				0.98 (0.003)

Intersection Similarity

Table 45 summarizes the intersection similarity measures between the four survival video games with an average of 0.7 (0.06).

Table 45. Intersection over union measures between survival video games (frequencies)

Intersection over Union	DST	ASTN	GRND	AVG (SE)
SUR	0.63	0.8	0.87	
DST		0.54	0.57	
ANT			0.78	
				0.7 (0.06)

Table 46 summarizes similarity measures between the video games within survival genre.

Table 46. Similarity measures' averages within survival genre

Similarity Values	**Percentages**	**Frequency**
Cosine Similarity	0.98 (0.003)	0.98 (0.003)
Intersection over Union	0.82 (0.01)	0.70 (0.06)

Comparisons Between Genres

Results in this section aim to assess the competency percentages' differences when compared between genres. The analysis aims to understand whether different cooperative genres induce different teaming competencies distributions.

Patterns of Teaming Competencies: Between Genres

Figure 17 shows line charts, each representing the average percentages of teaming competencies and standard error for an entire genre.

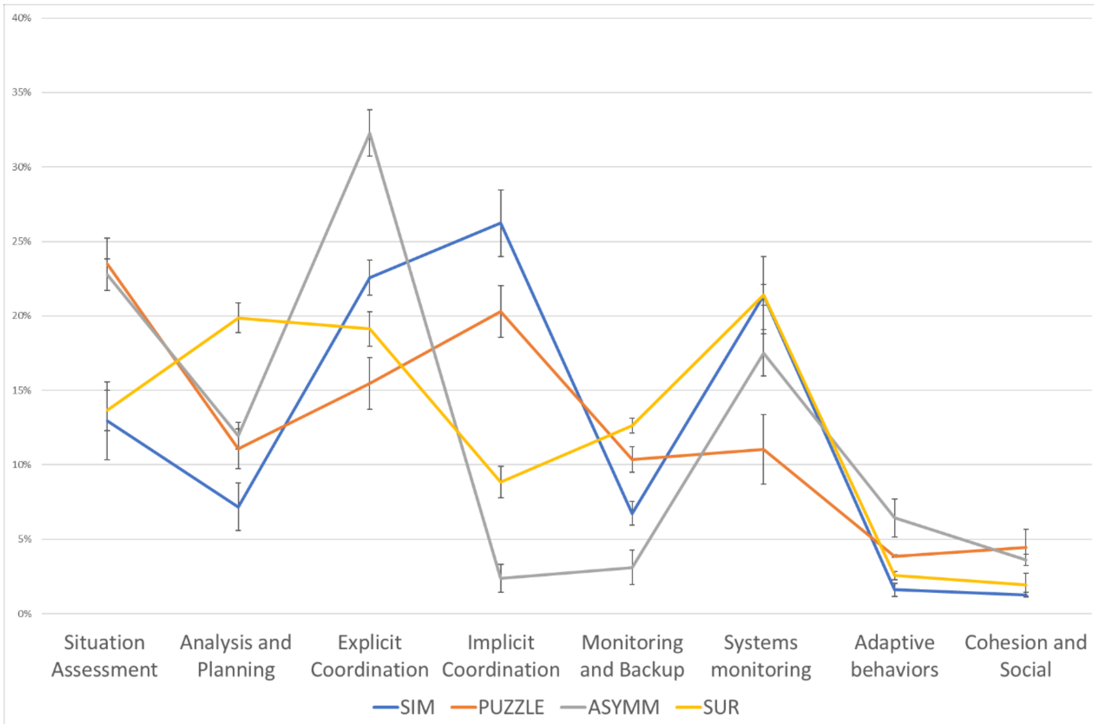


Figure 17. Percentages of teaming competencies for every genre

Table 47 summarizes the average percentages and standard error within every genre. The results are reported as AVG (SE).

Table 47. Average percentages of teaming competencies for all genres

Competency	Simulation	Puzzle Platformer	Asymmetric	Survival
Situation Assessment	13.0 (2.6)	23.5 (2.0)	22.8 (1.1)	13.7 (1.6)
Analysis and Planning	7.2 (1.6)	11.1 (1.5)	11.9 (0.9)	19.9 (1.2)
Explicit Coordination	22.6 (1.2)	15.5 (1.9)	32.3 (1.6)	19.1 (1.3)
Implicit Coordination	26.2 (2.2)	20.3 (1.9)	2.4 (0.9)	8.8 (1.2)
Monitoring and Backup	6.7 (0.8)	10.4 (1.0)	3.1 (1.2)	12.6 (0.6)
System Monitoring	21.4 (2.6)	11.0 (2.6)	17.5 (1.6)	21.4 (0.8)
Adaptive Behaviors	1.3 (0.2)	3.9 (0.1)	6.4 (1.3)	2.6 (0.3)
Cohesion and Social	1.6 (0.4)	4.5 (1.4)	3.6 (0.4)	1.9 (0.9)

Chi square test of homogeneity

A Chi Square of homogeneity was run to compare the distributions of the competencies between genres, with a X^2 (21) = 381.930 (p<.001) with Cramer's v equal to 0.217 which is a moderate effect size. A post hoc comparisons of proportions was conducted with Bonferroni correction. Significant differences were observed for every competency (see letters report in Table 48).

Table 48. Post-hoc pairwise comparisons between genres

		Video Game Genres			
		ASYMM	PUZZLE	SIM	SUR
Teaming Competency	SA	A	A	B	B
	AP	A	AB	B	C
	EC	A	B	C	BC
	IC	A	B	B	C
	MB	A	B	C	B
	SM	A	B	A	A
	AB	A	A	B	AB
	CS	AB	B	A	AB

Cosine Similarity

Table 49 shows the average cosine similarity and standard error values for video game pairwise comparisons between genres.

Table 49. Cosine similarity measures between genres (percentages)

Cosine Similarity	**Simulation**	**Asymmetric**	**Survival**	**AVG (SE)**
Puzzle Platformer	0.87 (0.01)	0.79 (0.01)	0.84 (0.01)	
Simulation		0.77 (0.01)	0.84 (0.01)	
Asymmetric			0.86 (0.01)	
				0.83(0.01)

Intersection Similarity

Table 50 shows the average intersection similarity and standard error values for video games pairwise comparisons between genres.

Table 50. Intersection over union similarity measures between genres (percentages)

Intersection	**Simulation**	**Asymmetric**	**Survival**	**Avg (SE)**
Puzzle Platformer	0.6 (0.02)	0.54 (0.01)	0.59 (0.01)	
Simulation		0.51 (0.02)	0.59 (0.01)	
Asymmetric			0.56 (0.01)	
				0.57(0.007)

Finally, Table 51 summarizes the average similarity measures within and between genres.

Table 51. Average similarity measures between genres

	Within Genre				
	Simulation	Platformer	Asymmetric	Survival	**Between**
Cosine (%)	0.94 (0.01)	0.94 (0.01)	0.98 (0.00)	0.98 (0.003)	0.82 (0.006)
Intersection (%)	0.73 (0.02)	0.75 (0.02)	0.8 (0.009)	0.82 (0.01)	0.57 (0.007)

Team Behaviors Frequency Between Genres

Figure 18 presents the average frequencies of teamwork behaviors for every teaming competency, for all four genres.

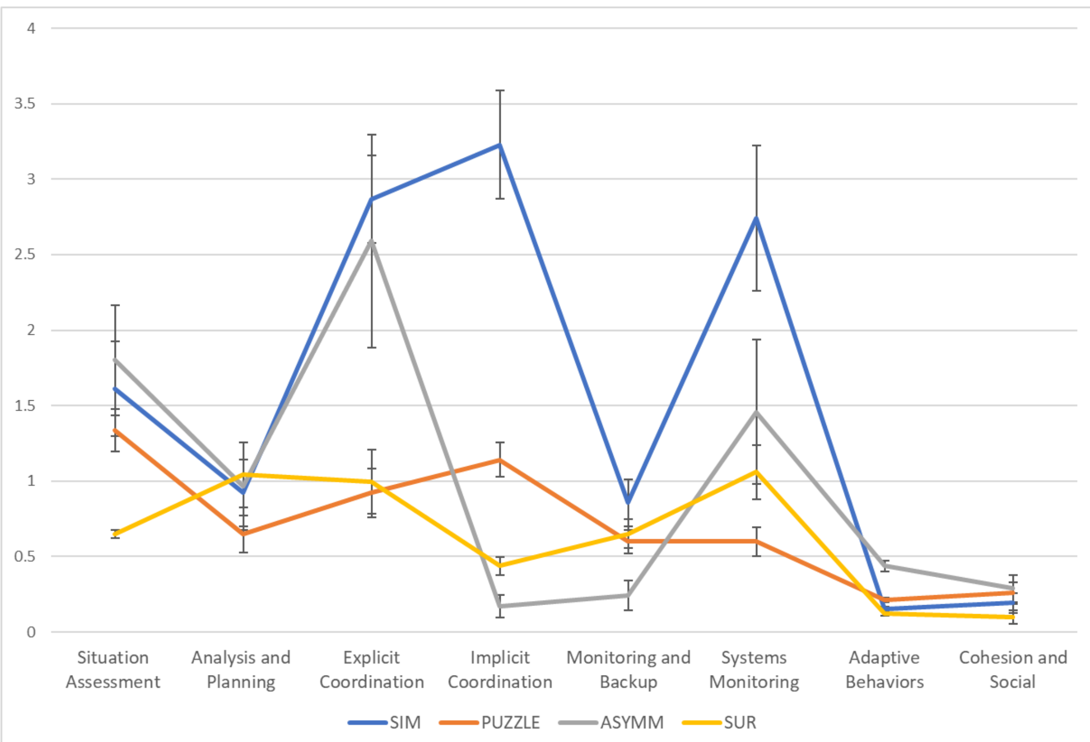


Figure 18. Frequencies of teaming competencies for all genres

Table 52 summarizes the average counts per minute and standard error within all four genres: Simulation (SIM), Puzzle Platformer (PUZZLE), Asymmetric (ASYMM) and Survival (SUR).

Table 52. Average frequencies of teaming competencies for all genres

Competency	**SIM**	**PUZZLE**	**ASYMM**	**SUR**
Situation Assessment	1.61 (0.31)	1.34 (0.14)	1.8 (0.36)	0.65 (0.03)
Analysis and Planning	0.92 (0.22)	0.65 (0.12)	0.97 (0.29)	1.04 (0.22)
Explicit Coordination	2.87 (0.29)	0.92 (0.16)	2.59 (0.71)	1.00 (0.21)
Implicit Coordination	3.23 (0.36)	1.14 (0.11)	0.17 (0.07)	0.44 (0.06)
Monitoring and Backup	0.86 (0.15)	0.6 (0.08)	0.24 (0.1)	0.65 (0.1)
Systems Monitoring	2.74 (0.48)	0.6 (0.1)	1.46 (0.48)	1.06 (0.18)
Adaptive Behaviors	0.15 (0.02)	0.21 (0.02)	0.44 (0.04)	0.13 (0.02)
Cohesion and Social	0.19 (0.07)	0.26 (0.07)	0.29 (0.09)	0.1 (0.04)

Cosine Similarity

Table 53 summarizes the cosine similarity values when compared between genres.

Table 53. Cosine similarity measures between genres (frequency)

Cosine Similarity	**Simulation**	**Asymmetric**	**Survival**	**Avg (SE)**
Puzzle Platformer	0.87 (0.01)	0.80 (0.01)	0.84 (0.01)	
Simulation		0.78 (0.02)	0.84 (0.01)	
Asymmetric			0.86 (0.01)	
				0.83 (0.01)

Intersection Similarity

Table 54 summarizes the intersection similarity values when compared between genres.

Table 54. Intersection over union similarity measures between genres (frequency)

Intersection Similarity	**Simulation**	**Asymmetric**	**Survival**	**Avg (SE)**
Puzzle Platformer	0.41 (0.01)	0.47 (0.03)	0.55 (0.01)	
Simulation		0.44 (0.02)	0.38 (0.02)	
Asymmetric			0.46 (0.05)	
				0.45 (0.01)

Table 55 summarizes the similarity measures within and between genres, with between genres scoring a lower average, of 0.83 (0.01) for cosine, and 0.45 (0.01) for intersection.

Table 55. Summary of similarity measures within and between genres

	Within Genre				
	Simulation	**Platformer**	**Asymmetric**	**Survival**	**Between**
Cosine	0.95 (0.005)	0.94 (0.02)	0.98 (0.002)	0.98 (0.003)	0.83 (0.01)
Intersection	0.63 (0.03)	0.68 (0.03)	0.54 (0.063)	0.7 (0.06)	0.45 (0.01)

Statistical Tests to Compare Similarity Measures

This paragraph presents the t-test results run to compare similarity values within and between genres. Test statistics and p-values are summarized in Table 56. All similarity measures were significantly different when comparing within measures to between measures. However, Cohen's d values were below 0.2, therefore suggesting no practical effect.

Table 56. Test statistics and p-values comparing similarity measures within and between genres

Similarity	Group 1	Group 2	t-test	*d*
Cosine (Percentages)	Within	Between	t (132.89) = 17.14 (*p*<.001***)	0.06
Intersection (Percentages)	Within	Between	t (151) = 13.49 (*p*<.001***)	0.07
Cosine (Frequency)	Within	Between	t (134.492) = 17.94 (*p*<.001***)	0.06
Intersection (Frequency)	Within	Between	t (37.867) = 3.484 (p<.001***)	0.13

RQ1 Discussion: Patterns of Teamwork Competencies

The distribution of teamwork competencies was investigated within and between genres. The following discussion addresses the implications of the findings, and propses the use of cooperative genres to induce targeted teamwork profiles based on the similarity of the teamwork distributions within genres. Furthermore, the discussion reflects on the compatibility of the results with teamwork and gaming literature, focusing on the cyclical and temporal nature of teamwork patterns, and on player modeling as an approach to predict players' actions and design accordingly. Through these findings, the work proposes using the cooperative genres as a design approach to target the associated teamwork profile in cooperative teamwork testbeds.

Using Cooperative Genres to Induce Targeted Teamwork Profiles

Studying patterns of behaviors is a prevalent theme in both teamwork and gaming literature. By understanding behavioral patterns, designers and researchers can predict individuals' behaviors and design the environment accordingly. The first research question in this study aimed to assess the similarity of distributions of teamwork competencies within video game cooperative genres and how they differ between genres. By showing similarities within genres, the study provides clarity on how teamwork behaviors are distributed in cooperative gaming environments, and how they are influenced by the cooperative design. Teamwork profiles were assessed to be homogeneous within genres, for most teaming competencies, where chi-square test showed no significant differences for seven out of eight competencies within puzzle platformer, and asymmetric genres, and six out of eight competencies within simulation. No significant differences within survival genre were observed. Therefore, the results support homogeneity of teaming competencies distributions within genres. Hence, the distribution of competencies within genres provides a genre-specific teamwork profile that can be targeted by testbed designers through t cooperative game features within a genre. Furthermore, since the findings are derived from several video games within every genre (e.g., five platformer games and a total of 50 teams), they are considered representative of the platforming genre, and therefore can be further generalized to using this genre in designing teamwork assessment testbeds. For example, puzzle platformer had the top three competencies of situation assessment, explicit and implicit coordination, and therefore the genre can be used to emphasize these competencies for assessment and training purposes. Expanding beyond the cooperative gaming context, the platforming genre can be perceived to replicate several real-world teamwork triggers, including information cues, and the need for collaborative solving and information exchange through interdependent tasks. An example would be collaborative learning environments, where cooperative games such as Portal 2 have been used to draw parallelism between the gaming environment and the learning environment (Yildirim, 2013). For example, students in collaborative learning need to perceive the problem cues and establish a common problem space through exchanging information and elaborating on cue meaning. Alternatively, while the asymmetric genre also emphasized situation assessment (22%), it had low implicit coordination percentages. Therefore, the top competencies to be induced through this genre are assessment, explicit coordination and systems monitoring. In real-world contexts, teams engage in a variety of asymmetric tasks, where team members have access to different types of information and roles. For example, in cockpit teamwork, pilots and air traffic controllers have access to asymmetric interfaces and rely on verbal communication to establish the team situation model, where the air traffic controllers need to coordinate with the crew, to direct the flight path, and monitor the environment such as reporting weather conditions and detecting risks like thunderstorms (Brannick et al., 2005). In contrast, simulation genre can be used to induce high percentages of implicit coordination, followed by explicit coordination and systems monitoring. This genre can parallel some aspects of real-world teamwork, where existing routine tasks are established, however they are applied in dynamic and unexpected situations under time pressure, such as emergency medicine, where delays in execution can have negative consequences (Fernandez et al., 2008). In emergency medicine situations, the medical staff have a variety of allocated roles (task allocation), with well-established processes, therefore staff members are capable of implicitly executing routine actions. Furthermore, while mission analysis and planning are necessary, they are limited while in action, and must be fast and brief (Fernandez et al., 2008). Finally, the survival genre had top competencies of analysis and planning, explicit coordination and systems monitoring. The survival genre can be paralleled to low pressure and longitudinal teamwork tasks, that can be executed over long periods of time, with a wide teamwork environment to monitor, such as organizational teamwork. In organizational teamwork, team members can execute a variety of tasks individually (emergent interdependence rather than mandatory) but must maintain coordination and communication with team members to report status updates (Salas et al., 2015). Furthermore, team members engage in regular meetings to maintain the communal nature of the work and engage in extensive planning and analysis (Salas et al., 2015). Other competencies also varied significantly between genres, such as monitoring and backup. Hence, every genre can be used to emphasize different competencies, and some cooperative gaming elements can be paralleled to real life situations for further generalization. However, this study did not empirically address the applicability of the findings beyond cooperative gaming and testbeds, and therefore future work can address the transfer effect of teamwork skills beyond cooperative games to similar real-world situations.

Significant differences were observed when comparing distributions between genres for all teaming competencies with $p<.001$ and an average Cramer's v of 0.217, equivalent to a moderate effect size. Therefore, supporting that the distribution of teamwork competencies is different between genres. The post hoc analysis compared each teamwork competency between all four genres, also showing significant differences. For example, asymmetric and simulation genres had six out of eight significant differences, with only systems monitoring and cohesion and social being non-significant. Similarly, puzzle platformer and simulation had six out of eight significantly different competencies, and only had two non-significant differences for analysis and planning and implicit coordination. While other genre pairwise comparisons had higher rates of non-significant competencies, such as asymmetric and puzzle platformer with four out of eight non-significant differences, and simulation and survival with five out of eight non-significant differences. Therefore, different cooperative genres require teams to distribute their behaviors differently, with some genres being more different than others. This analysis suggests the use of cooperative genre to target different distributions of teamwork competencies.

Exploring Cyclical and Temporal Patterns

Teamwork literature suggests that teams follow patterns of teaming competencies. Specifically, cyclical patterns of teaming behaviors have been frequently proposed. Examples include the four phases of preparation, execution, evaluation, and adjustment, suggested as an integrated framework of teaming behaviors (Rousseau et al., 2006) , or the cyclical pattern of transition and execution processes (Marks et al., 2001) . Other frameworks include teams' adaptation patterns of situation assessment, plan formulation, plan execution and team learning (Rosen et al., 2011) . These frameworks suggest that teams follow a pattern of behavior. These behaviors generally follow specific orders and happen in repeated cycles. Additionally, behaviors under every category can differ in frequency and timing, depending on the environment (Marks et al., 2001) . In this study, behavioral markers were grouped under teaming competencies. The results support that within the same cooperative genre, these behaviors are similarly distributed. In contrast the distribution was significantly different between genres. This suggests that different solution strategies are being utilized with different teamwork patterns between genres.

Player Modeling as a Predictive Approach to Designing Gamified Testbeds

In gaming literature, player modeling is a common method, that aims to understand players' behaviors, actions, and intentions (Bakkes et al., 2012; Hooshyar et al., 2018). Modeling player actions, tactics and strategies can follow several data-driven approaches, including in-game observations of behaviors, automatic data generated from the game, subjective data, and objective biometric data (Bakkes et al., 2012; Hooshyar et al., 2018) . Player modeling reveals patterns of behaviors, associated with game states and conditions, allowing designers and researchers to predict players behaviors based on the game state, and adapt the games' conditions following player's needs.

Therefore, in the light of the patterns theme, the similarity analysis conducted in this study has several implications. First, a pattern of teaming competencies' distributions was observed within genres, indicated by the insignificant differences from the chi-square test, and the high similarity measures. From a team science perspective, this aligns with teamwork research that focuses on patterns of teamwork behaviors, and provides empirical support for theoretical frameworks, that suggest different frequencies and times depending on the teaming environment. Moreover, it confirms that cooperative gaming environments are suitable environments to study teamwork since theoretical teamwork frameworks apply in these environments. From a gaming science perspective, the analysis represents a form of players modeling that targeted teams rather than individual players. The revealed patterns show that within genres, teams engage in similar distributions of competencies. Hence, designers and researchers aiming to use cooperative games to foster, train and assess teamwork, are provided with a predictive teamwork profile, with consistent competencies distributions. Cooperative genres can be chosen to produce the desired teamwork patterns.

CHAPTER 5. COOPERATIVE FEATURES-TEAMWORK BEHAVIORS ASSOCIATIONS (RQ2)

The next section of results aims to present the frequency of associations between the cooperative features and teaming competencies, to explore what induced the teaming competencies within genres. This section will present visual representations of these associations to illustrate how teaming competencies within genres were associated with the genre's cooperative features. The analysis aims to detect the most frequent associations, and provide them as design guidance, where cooperative features can be used to emphasize the behaviors with which it was most frequently associated.

First, the section presents area graphs, with a breakdown of every teaming competency into the three top cooperative features associated with it. This analysis aims to understand what cooperative features led to the competency distributions analyzed in the previous consistency analysis. Therefore, presenting the dominating cooperative features that drove the teaming behaviors in the genre.

Second, frequency of associations between cooperative features and teaming competencies will be presented through bubble graphs, and a breakdown of the top three competencies associated with every cooperative feature. This analysis aims to provide an understanding of the top teaming competencies induced by every cooperative feature, and how different cooperative features can be used to trigger different distributions of teaming behaviors. Higher percentages of associations means that the feature induced more of a specific behavior, therefore emphasizing the need for these top competencies when engaging with the specific cooperative feature and providing design guidance to use the features to emphasize the top teaming competencies associated with them.

Coding Remarks

Up to two features were coded in one annotation, when they simultaneously induced a behavior (e.g., SE/ER: SM-E: shared environment and environment resources, were associated with a system monitoring behavior). In these cases, two pairs would be counted, one accounting for SE: SM-E and one for ER: SM-E. In survival genre, two features and one dynamic can be coded in the same annotation. These dynamics are Team Spirit (TS) and Community Survival (CS). A code example would be LLR/SA (TS): PBB-R, which translates into limited life resources (LLR) and sharing abilities (SA), paired with a team spirit (TS) dynamic, were associated with a proactive backup behavior of sharing resources (PBB-R).

Aggregate counts of associations were generated in contingency tables for every genre, summing across all teams and video games within the genre. The aggregate sums and percentages used in this analysis section are not the aggregate sum of behaviors, but rather the count of times an association between a cooperative feature and a behavioral marker was coded.

Teaming Competencies and Cooperative Features: Breakdown Within Genres

Puzzle Platformer

Figure 19 illustrates an area graph of the percentages of cooperative feature-teaming competency within the puzzle platformer genre. An aggregate sum of cooperative feature-teamwork behavior was calculated for all games within puzzle platformer (five video games, 50 teams), and percentages of pairs were generated. The x-axis presents the teaming competencies, the y-axis the percentages, broken down into the inducing cooperative features. The legend of the cooperative features is presented to the left of the y-axis.

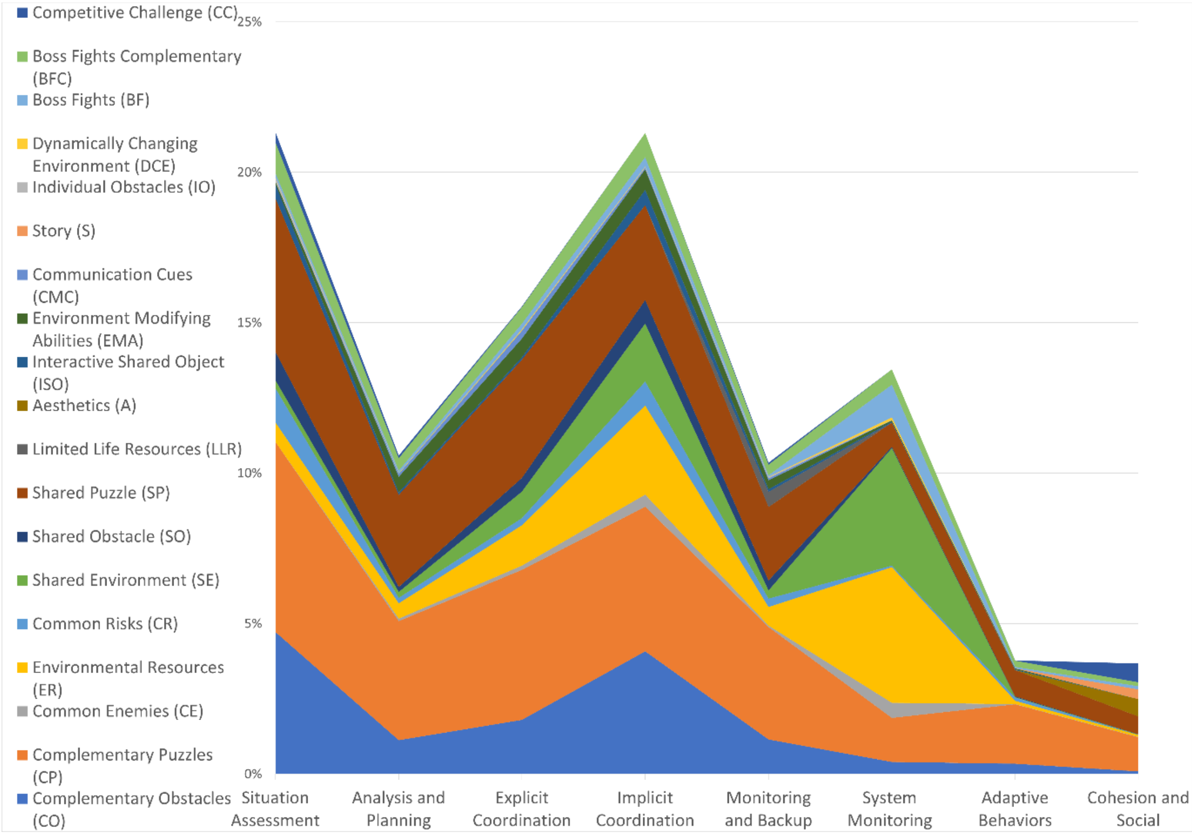


Figure 19. Area graph of puzzle platformer teaming competencies and cooperative features

Table 57 presents the top three percentages of cooperative features in every teaming competency. The area presented by CO within situation assessment in Figure 19, accounts for 22% of the total situation assessment, CP accounts for 30% and SP accounts for 24%.

Table 57. Top three cooperative features associated with every teaming competency in puzzle platformer

Teaming Competency	**Percentages of top three cooperative features**		
Situation Assessment	CO (22%)	CP (30%)	SP (24%)
Analysis and Planning	CP (37%)	SP (29%)	CO (11%)
Explicit Coordination	CP (32%)	SP (25%)	CO (12%)
Implicit Coordination	CP (23%)	CO (19%)	SP (15%)
Monitoring and Backup	CP (36%)	SP (24%)	CO (11%)
Systems Monitoring	ER (33%)	SE (29%)	CP (11%)
Adaptive Behaviors	CP (52%)	SP (23%)	CO (9%)
Cohesion and Social	CP (31%)	CC (17%)	SP, A (16%)

Asymmetric

Figure 20 illustrates an area graph of the percentages of cooperative feature-teaming competency within the Asymmetric genre.

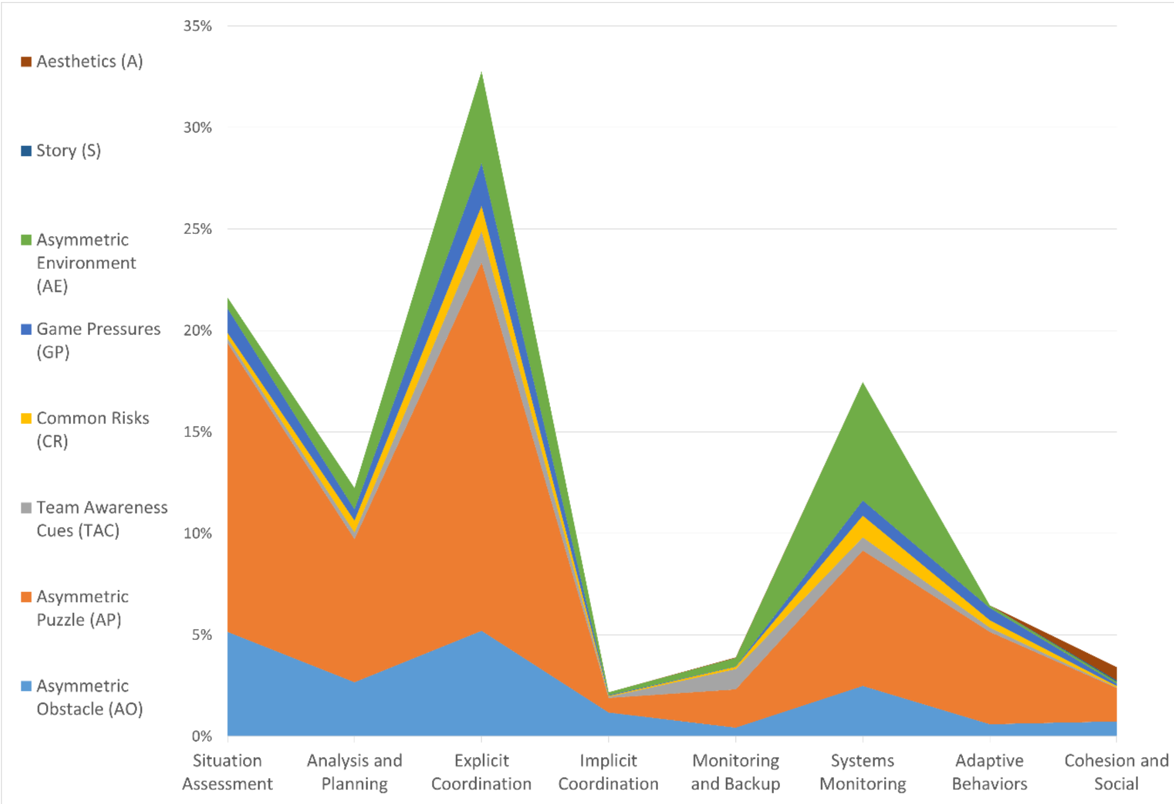


Figure 20. Area graph of asymmetric teaming competencies and cooperative features

Table 58 presents the top three percentages of cooperative features involved in every teaming competency within Asymmetric genre. For example, Asymmetric Puzzle (AP) accounts for 66% of the Situation Assessment behaviors in the asymmetric genre, followed by Asymmetric Obstacle (AO) accounting for 24%. Alternatively, AO accounts for 54% of implicit coordination.

Table 58. Top three cooperative features associated teaming competencies in asymmetric

Teaming Competency	**Percentages of top three cooperative features**		
Situation Assessment	AP (66%)	AO (24%)	GP (6%)
Analysis and Planning	AP (58%)	AO (22%)	AE (9%)
Explicit Coordination	AP (55%)	AO (16%)	AE (14%)
Implicit Coordination	AO (54%)	AP (33%)	AE (7%)
Monitoring and Backup	AP (48%)	TAC (26%)	AO, AE (11%)
Systems Monitoring	AP (38%)	AE (33%)	AO (14%)
Adaptive Behaviors	AP (70%)	AO, GP (9%)	CR (6%)
Cohesion and Social	AP (48%)	AO (22%)	A (19%)

Simulation

Figure 21 illustrates an area graph of the percentages of cooperative feature-teaming competency within the Simulation genre. An aggregate sum was calculated for all the pairs of cooperative feature-teaming competency within Simulation (five video games, 47 teams) and percentages of pairs were generated

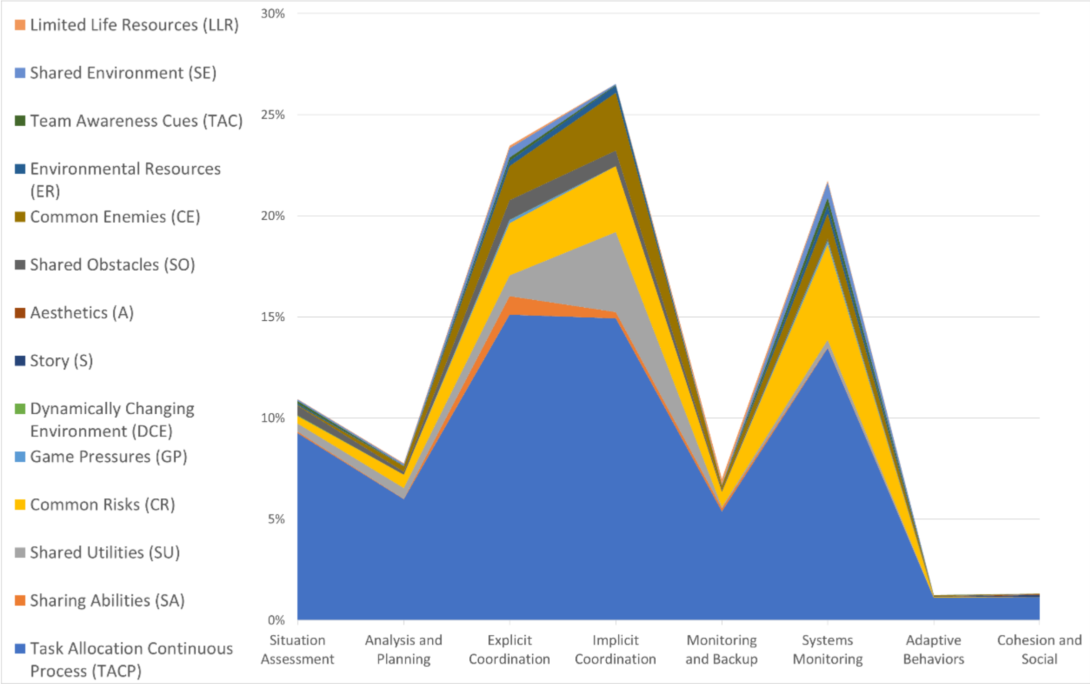


Figure 21. Area graph of simulation teaming competencies and cooperative features

Table 59 presents the top three percentages of cooperative features involved in every teaming competency within simulation genre.

Table 59. Top three cooperative features associated with teaming competencies in simulation

Teaming Competency	Percentages of top three cooperative features		
Situation Assessment	TACP (85%)	SU, SO (4%)	CR (3%)
Analysis and Planning	TACP (77%)	CR (9%)	SU (7%)
Explicit Coordination	TACP (64%)	CR (11%)	CE (7%)
Implicit Coordination	TACP (56%)	SU (15%)	CR (12%)
Monitoring and Backup	TACP (77%)	CR (11%)	LLR (4%)
Systems Monitoring	TACP (62%)	CR (22%)	CE (6%)
Adaptive Behaviors	TACP (89%)	SO, CE, CR (3%)	SA, TAC (1%)
Cohesion and Social	TACP (95%)	SO, CE (2%)	SA (1%)

Survival

Figure 22 illustrates an area graph of the percentages of cooperative feature-teaming competency within the survival genre.

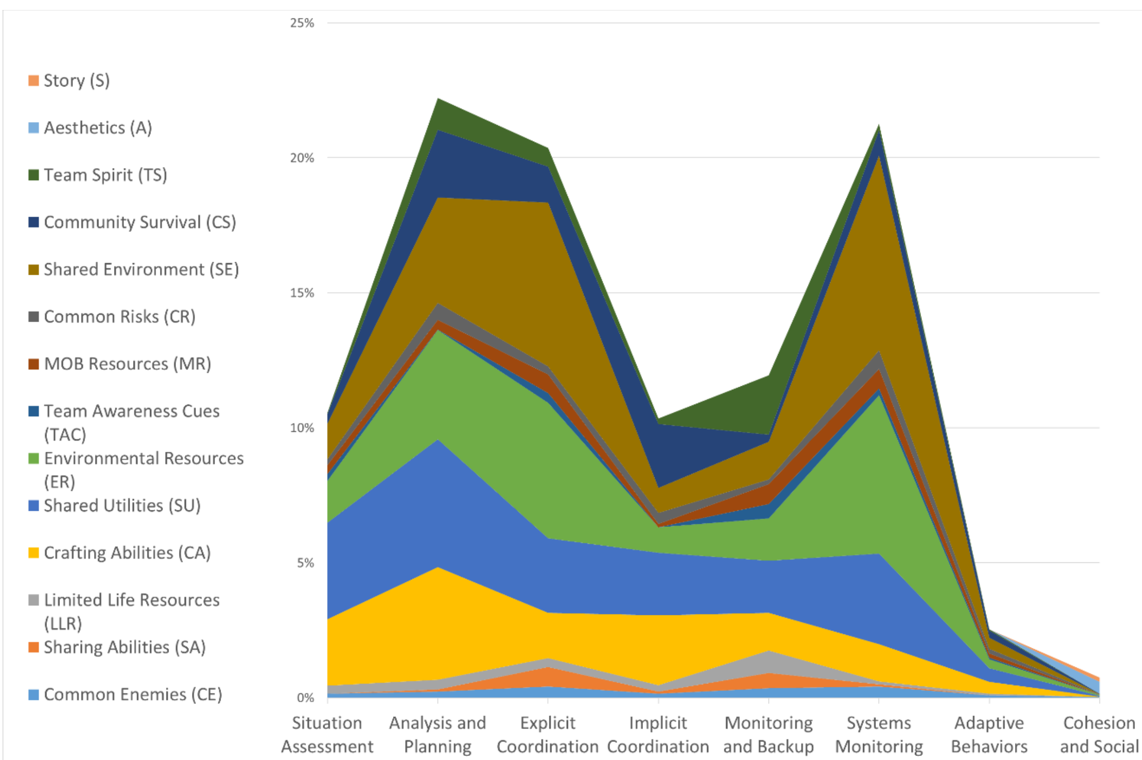


Figure 22. Area graph of survival teaming competencies and cooperative features

Table 60 presents the top three percentages of cooperative features involved in every teaming competency within survival genre.

Table 60. Top three cooperative features associated with teaming competencies in survival

Teaming Competency	Percentages of top three cooperative features		
Situation Assessment	SU (34%)	CA (23%)	ER (15%)
Analysis and Planning	SU (21%)	CA (19%)	ER (18%)
Explicit Coordination	SE (30%)	ER (25%)	SU (14%)
Implicit Coordination	CA (25%)	CS, SU (23%)	ER, SE (9%)
Monitoring and Backup	TS (18%)	SU (16%)	ER (13%)
Systems Monitoring	SE (34%)	ER (28%)	SU (16%)
Adaptive Behaviors	SU (20%)	CA (18%)	SE (10%)
Cohesion and Social	A (59%)	S (20%)	CE, ER (5%)

Counts of Cooperative Feature-Competency Pairs: Bubble Graphs

This analysis aims to explore the associations between cooperative features and teamwork behaviors. By establishing the most frequent associations, every cooperative feature is presented with the top three competencies associated with it. This analysis aims to establish a clearer relationship between the cooperative features and teamwork behaviors and provide an empirical foundation for further design recommendations on how to use features to design teamwork cooperative testbeds.

Puzzle-Platformer

Figure 23 is a bubble chart of cooperative feature-behavioral markers pairs within the puzzle platformer genre. The values are aggregated for all five video games (50 teams). The x axis represents the cooperative features observed in the genre, and the y axis the teaming competencies. Every bubble is an aggregate sum of coded pairs across all video games within puzzle platformer.

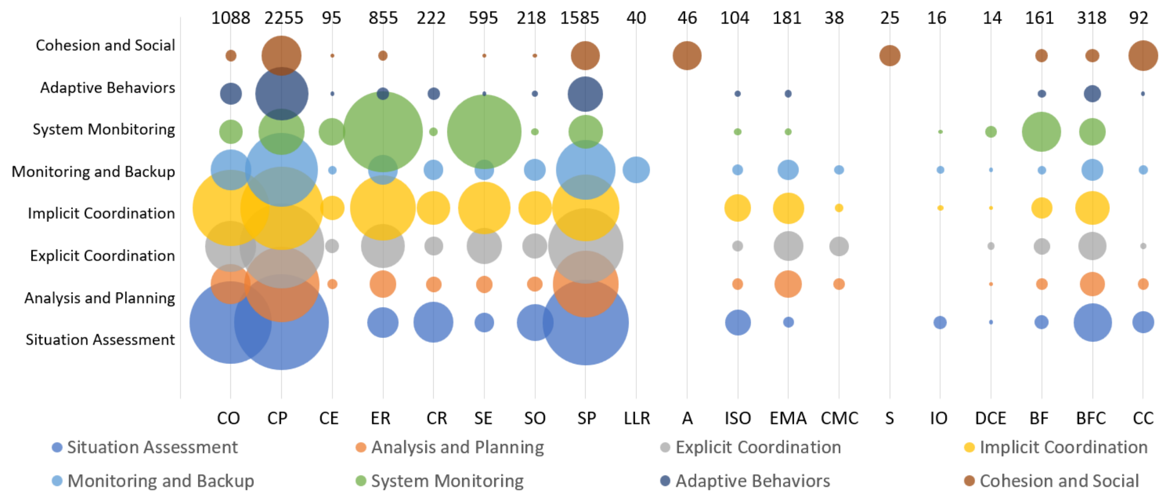


Figure 23. Associations between platformer cooperative features and teamwork competencies

Table 61 summarizes the top three teamwork competencies associated with every cooperative features within puzzle platformer. For example, 34% of behaviors associated with complementary obstacle were situation assessment (SA).

Table 61. Top three teamwork competencies associated with cooperative features in puzzle platformer

Cooperative Feature	**Top Three Teamwork Competencies**		
Complementary Obstacle (CO)	SA (34%)	IC (30%)	EC (13%)
Complementary Puzzle (CP)	SA (22%)	EC, IC (18%)	AP, MB (13%)
Common Enemies (CE)	SM (42%)	IC (34%)	EC (12%)
Environmental Resources (ER)	SM (42%)	IC (27%)	EC (13%)
Common Risks (CR)	SA (41%)	IC (29%)	MB (10%)
Shared Environment (SE)	SM (52%)	IC (26%)	EC (12%)
Shared Obstacle (SO)	SA (34%)	IC (28%)	EC (17%)
Shared Puzzle (SP)	SA (26%)	EC (20%)	IC, AP (15%)
Limited Life Resources (LLR)	MB(100%)		
Aesthetics (A)	CS (100%)		
Interactive Shared Object (ISO)	IC (39%)	SA (36%)	AP, EC, MB (7%)
Environment Modify Abilities (EMA)	IC (30%)	EC (27%)	AP (23%)
Communication Cues (CMC)	EC (55%)	AP (21%)	MB (13%)
Story (S)	CS (100%)		
Individual Obstacles (IO)	SA (63%)	MB (19%)	IC (13%)
Dynamic Changing Environment(DCE)	SM (50%)	EC (21%)	SA, AP, IC, MB(7%)
Boss Fight (BF)	SM (53%)	IC (16%)	EC (9%)
Boss Fight Challenge (BFC)	SA (26%)	IC (20%)	EC, SM (13%)
Competitive Challenge (CC)	CS (53%)	SA (29%)	AP (8%)

Asymmetric

Figure 24 is a bubble chart of cooperative feature-behavioral markers pairs within the asymmetric genre.

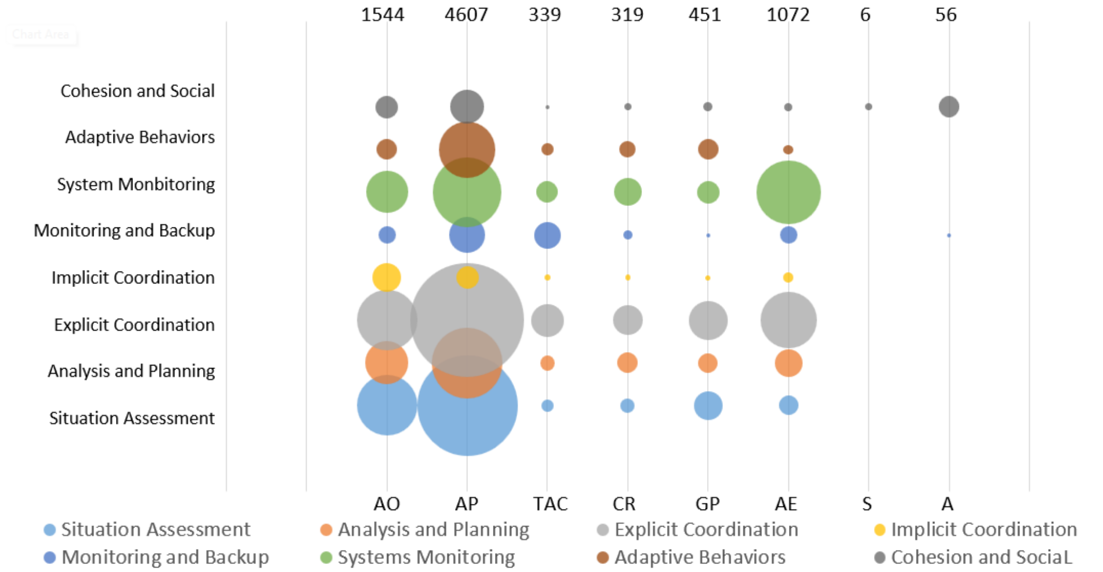


Figure 24. Associations between asymmetric cooperative features and teamwork competencies

Table 62 summarizes the top three teamwork competencies associated with every cooperative feature with asymmetric genre.

Table 62. Top three teamwork competencies associated cooperative features in asymmetric

Cooperative Feature	**Top Three Competencies**		
Asymmetric Obstacle (AO)	SA (28%)	EC (28%)	AP, SM (13%)
Asymmetric Puzzle (AP)	EC (33%)	SA (26%)	AP, SM (13%)
Team Awareness Cues (TAC)	EC (38%)	MB (25%)	SM (16%)
Common Risks (CR)	EC (33%)	SM (28%)	AP (15%)
Game Pressures (GP)	EC (39%)	SA (22%)	SM (14%)
Asymmetric Environment (AE)	SM (46%)	EC (35%)	AP (8%)
Story (S)	CS (100%)		
Aesthetics (A)	CS (96%)	MB (4%)	

Simulation

Figure 25 is a bubble chart of cooperative feature-behavioral markers pairs within the simulation genre.

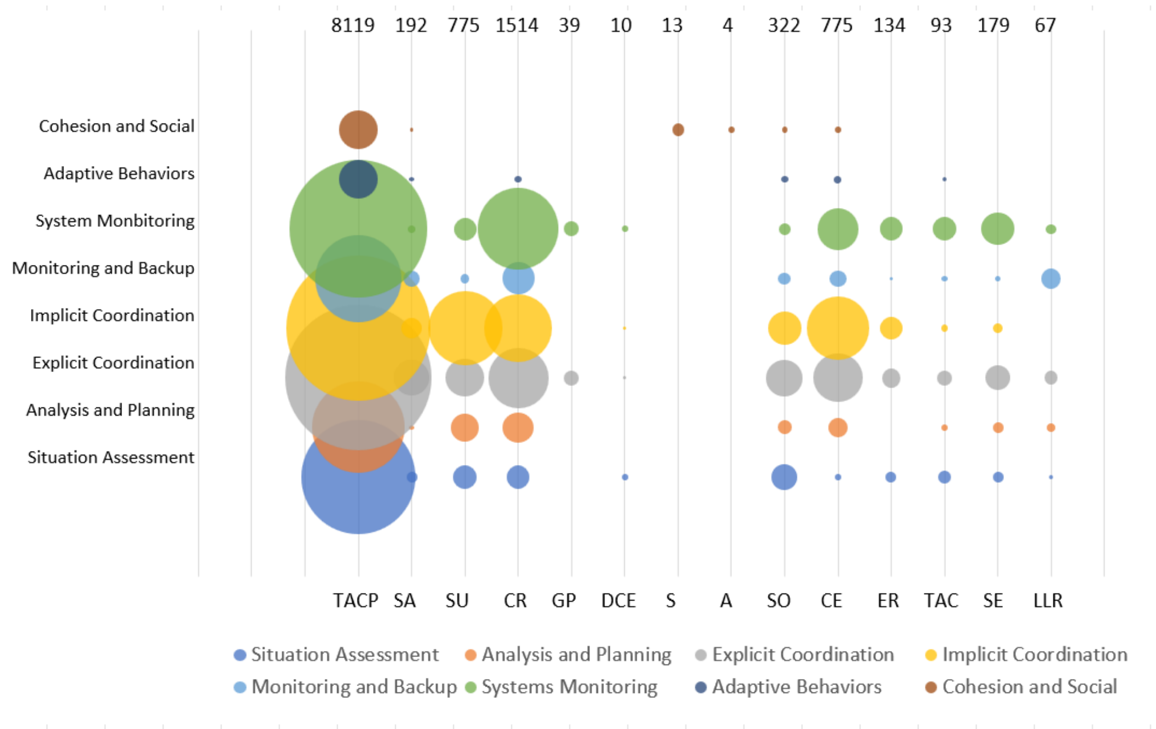


Figure 25. Associations between simulation cooperative features and teamwork competencies

Table 63 summarizes the top three teamwork competencies associated with every cooperative feature with simulation genre. For example, 23% of behaviors associated with task allocation (TACP) were explicit coordination (EC), 22% implicit coordination (IC) and 20% systems monitoring (SM), implying that using this cooperative feature in testbed design can emphasize these three teamwork competencies.

Table 63. Top three teamwork competencies associated cooperative features in simulation

Cooperative Feature	Top Three Competencies		
Task Allocation Continuous Process (TACP)	EC (23%)	IC (22%)	SM (20%)
Sharing Ability (SA)	EC (58%)	IC (21%)	MB (11%)
Shared Utility (SU)	IC (62%)	EC (16%)	AP (9%)
Common Risks (CR)	SM (38%)	IC (26%)	EC (21%)
Game Pressure (GP)	SM (49%)	EC (51%)	
Dynamically Changing Environment (DCE)	SM, SA (40%)	EC, IC (10%)	
Story (S)	CS (100%)		
Aesthetics (A)	CS (100%)		
Shared Obstacle (SO)	EC (37%)	SA (18%)	IC (29%)
Common Enemy (CE)	IC (45%)	EC (27%)	SM (19%)
Environmental Resources (ER)	SM (36%)	IC (32%)	EC (24%)
Team Awareness Cues (TAC)	SM (52%)	EC (31%)	SA,AP (6%)
Shared Environment (SE)	SM (52%)	EC (31%)	SA,AP(6%)
Limited Life Resources (LLR)	MB (51%)	SM (13%)	EC (24%)

Survival

Figure 26 is a bubble chart of cooperative feature-behavioral markers pairs within the survival genre.

Figure 26. Associations between survival cooperative features and teamwork competencies

Table 64 summarizes the top three teamwork competencies associated with every cooperative feature with survival genre.

Table 64. Top three teamwork competencies associated with cooperative features in survival

Cooperative Feature	Top Three Competencies		
Common Enemies (CE)	SM (23%)	EC (23%)	MB (19%)
Sharing Abilities (SA)	EC (48%)	MB (38%)	IC (6%)
Limited Life Resources (LLR)	MB (37%)	AP, EC (15%)	SA (13%)
Crafting Abilities (CA)	AP (30%)	IC, SA (17%)	EC (12%)
Shared Utilities (SU)	AP (25%)	SA (19%)	SM (18%)
Environmental Resources (ER)	SM (30%)	EC (26%)	AP (21%)
Team Awareness Cues (TAC)	MB (36%)	EC (24%)	SA, SM (17%)
MOB Resources (MR)	MB (24%)	SM (23%)	EC (22%)
Common Risks (CR)	SM (26%)	AP (24%)	IC (16%)
Shared Environment (SE)	SM (34%)	EC (29%)	AP (18%)
Community Survival (CS)	AP (31%)	IC (29%)	EC (17%)
Team Spirit (TS)	MB (48%)	AP (25%)	IC (15%)
Aesthetics (A)	CS (100%)		
Story (S)	CS (100%)		

RQ2 Discussion: Associations Between Cooperative Features and Teamwork Behaviors

Associations between cooperative features and teamwork competencies were investigated. The following discussion aims to propose associations as design tools, through using the cooperative features to emphasize the teamwork competencies most frequently associated with them.

Associations as Testbed Design Guidance

There has been interest in previous work to understand the associations between cooperative game features and social interactions, including teamwork (Marlow et al., 2016; Morchheuser et al., 2017; Hämäläinen et al., 2018). Particularly it has been suggested that game mechanics can be designed in accordance with the targeted activity. Previous work aimed to study how game mechanics can trigger collaboration and social interactions, by manipulating interdependence levels (Harris et al., 2016), game mechanics (Hämäläinen et al., 2018) and design patterns (Emmerich & Masuch, 2017). Moreover, Marlow et al., (2016) suggested a theoretical research plan, with a prerequisite assumption that there is a relationship between game attributes and teamwork behaviors. This work provides empirical data that reveals associations between cooperative features and teaming behaviors, and aims to answer the second research question, asking about the associations between features and behaviors. Through understanding these associations, the work provides design guidance, to use the cooperative features to emphasize the most frequently associated behaviors.

Understanding the Underlying Features inducing Competencies within Genres

First, by breaking down the teamwork profiles analyzed in the first research question, into the underlying cooperative features (through the area graphs and top percentages), it was observed that in every genre, the dominant cooperative features induced most teamwork behaviors. Complementary puzzles ranked as the top cooperative feature inducing analysis and planning, explicit coordination, implicit coordination, monitoring and backup and adaptive behaviors, in the puzzle platformer genre. Similarly, asymmetric puzzles in the asymmetric genre dominated most of the competencies, contributing to up to 66% of the situation assessment behaviors in this genre. The task allocation continuous process also dominated the induced competencies in simulation genre. More variety was observed in the survival genre, where shared utilities, shared environment, crafting abilities and team spirit were observed as the top one features across the competencies. This analysis aimed to explain what underlying features are driving the teamwork profiles and aims to answer “why” these cooperative genres induce teamwork behaviors.

Using Cooperative Features to Emphasize Teamwork Competencies Through Most Frequent Associations

Associations between cooperative features and teamwork behaviors were derived through bubble charts, and the top three competencies induced by every cooperative feature were presented. This analysis aimed to answer the relationship between cooperative features and teamwork behaviors and provided empirically supported guidelines on what cooperative features to use to induce the desired teamwork distributions.

Alternatively, environmentally oriented features, such as shared environment, common enemies, and environmental resources, were most frequently associated with systems monitoring and coordination, across all genres, suggesting their use in situations where monitoring external and internal environmental conditions is desired. For example, puzzles in the puzzle platformer genre (complementary obstacles, complementary puzzles, shared puzzles, and shared obstacles), were most frequently associated with situation assessment, coordination (implicit and explicit) and analysis and planning, implying that these features can be used when designers are targeting transition processes that require cue recognition, cue meaning, and collaborative problem solving, with an emphasis on both implicit and explicit coordination. Similarly, in the asymmetric genre, situation assessment, explicit coordination and analysis and planning also were in the top competencies for the asymmetric puzzles, showing similarities in how puzzles across genres induce behaviors. However, in this case, asymmetric puzzles had a higher emphasis on explicit coordination, with small to no emphasis on implicit coordination. Implying that cooperative puzzles can be manipulated, by adapting a puzzle platformer approach (shared environment with interdependent tasks), or asymmetric approach (asymmetric environment with interdependent tasks), to place more emphasis on behaviors.

The design insights chapter (chapter 7) dives into more detail, elaborating on associations between features and behaviors. This analysis showed that different cooperative features have different distributions of teamwork behaviors and can be used to induce desired teamwork behaviors depending on the designers' and researchers' targeted constructs.

CHAPTER 6. COMPARING TEAMWORK BEHAVIORS BASED ON PERFORMANCE OUTCOMES (RQ3)

This chapter presents the results and discussion comparing upper and lower performers within every genre (RQ3). This analysis aims to explore the differences in the frequency of teamwork behaviors between upper and lower performers, to gain insights on what teamwork competencies affected the performance outcomes, and how these differences vary depending on the genre.

Puzzle Platformer

Figure 27 is a star chart showing the average frequency of teaming competencies per minute for the upper and lower half performers, within the puzzle platformer genre.

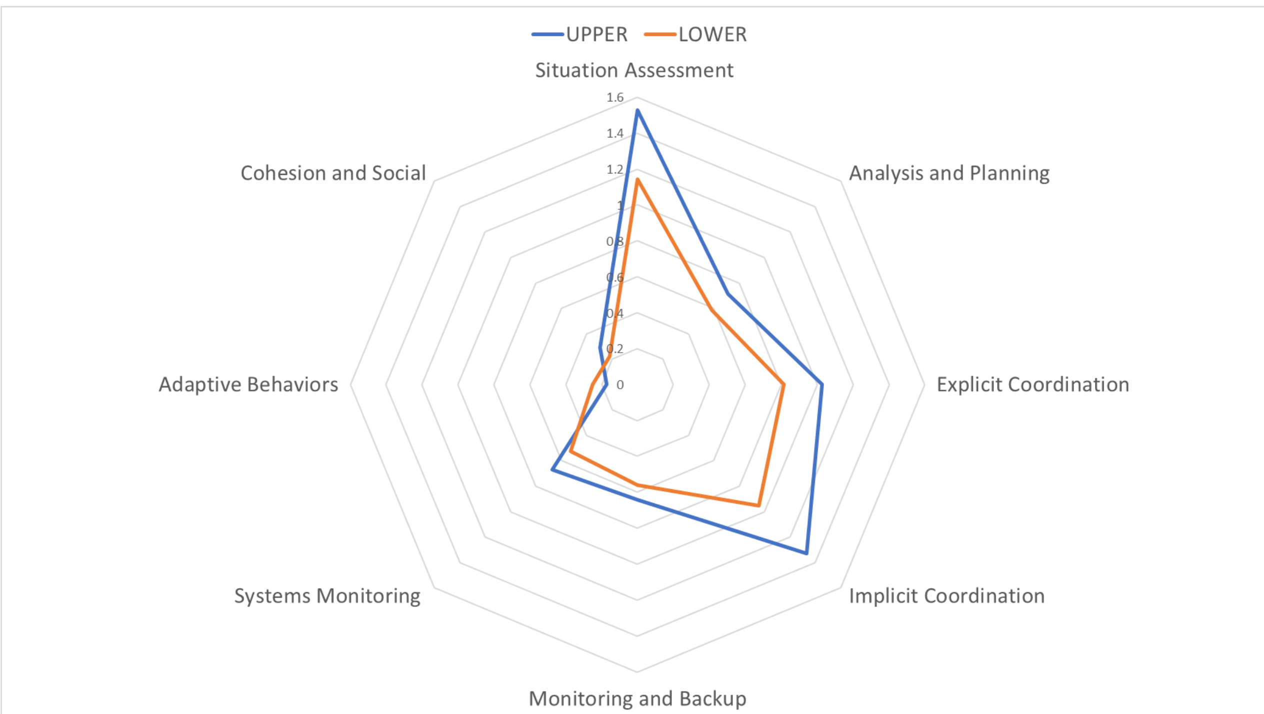


Figure 27. Frequency of teaming competencies between upper and lower performers in puzzle platformer genre

Table 65 presents the p values from the t-tests ran to compare every teaming competency between upper and lower half performers. Statistical significance with moderate effect size is observed for situation assessment (*p*=.005, Cohen's *d* = 0.47) and implicit coordination (*p*=.004, Cohen's *d* = 0.45). Statistical significance with small effect size is observed for adaptive behaviors (*p*=0.048, Cohen's *d*=0.13).

Table 65. t-test results comparing every teaming competency between upper and lower performers in Puzzle Platformer. P values lesser than 0.05 are indicated with one star (*), p values lesser than 0.01 with two stars (**).

Teaming Competency	**t-test (*p*-value)**	**Cohen's d**
Situation Assessment	t (42) = 2.93 (.005**)	0.47
Analysis and Planning	t (37) = 1.24 (.224)	
Explicit Coordination	t (48) = 1.59 (.118)	
Implicit Coordination	t (48) = 2.97 (.004**)	0.45
Monitoring and Backup	t (48) = 1.02 (.214)	
Systems Monitoring	t (48) = 1.99 (.0525)	0.25
Adaptive Behaviors	t (48) = -2.03 (.0476*)	0.13
Cohesion and Social	t (43) = 1.3 (.201)	

Asymmetric

Figure 28 is a star chart showing the average frequency of teaming competencies per minute for the upper and lower half performers, within the asymmetric genre. The upper and lower half performers were separated within every video game in asymmetric, based on their time to finish levels and number of errors. The statistical test compares the aggregate of upper and lower performers for all four video games within asymmetric.

Figure 28. Frequency of teaming competencies between upper and lower performers in asymmetric genre

Table 66 presents the test statistics and p values for every teaming competency between upper and lower performers. Statistical significance with large effect size is observed for situation assessment (*p*=.047, *d* = 0.81), and small effect size for cohesion and social (*p*=.02, *d*=0.19)

Table 66. t-test results comparing every teaming competency between upper and lower performers in Asymmetric. P values less than 0.05 are indicated with one star (*).

Teaming Competency	**t-test**	**Cohen's d**
Situation Assessment	t (36) = 2.05, *p* = .047*	0.81
Analysis and Planning	t (36) = 1.36, *p* = .182	
Explicit Coordination	t (36) = 1.21, *p* = .233	
Implicit Coordination	t (36) = 1.25, *p* = .218	
Monitoring and Backup	t (36) = 1.03, *p* = .306	
Systems Monitoring	t (36) = 1.37, *p* = .178	
Adaptive Behaviors	t (36) = -1.76, *p* = .08	
Cohesion and Social	t (36) = 2.31, *p* = .02*	0.19

Simulation

Figure 29 is a star chart showing the average frequency of teaming competencies per minute for the upper and lower half performers, within the simulation genre. Upper and lower performers were separated within every video game in the simulation genre, based on scores provided by the game (e.g., in Overcooked 2 the game provides the count of dishes delivered, count of errors and the total score). The test statistics compare the aggregate of upper and lower performers within simulation genre.

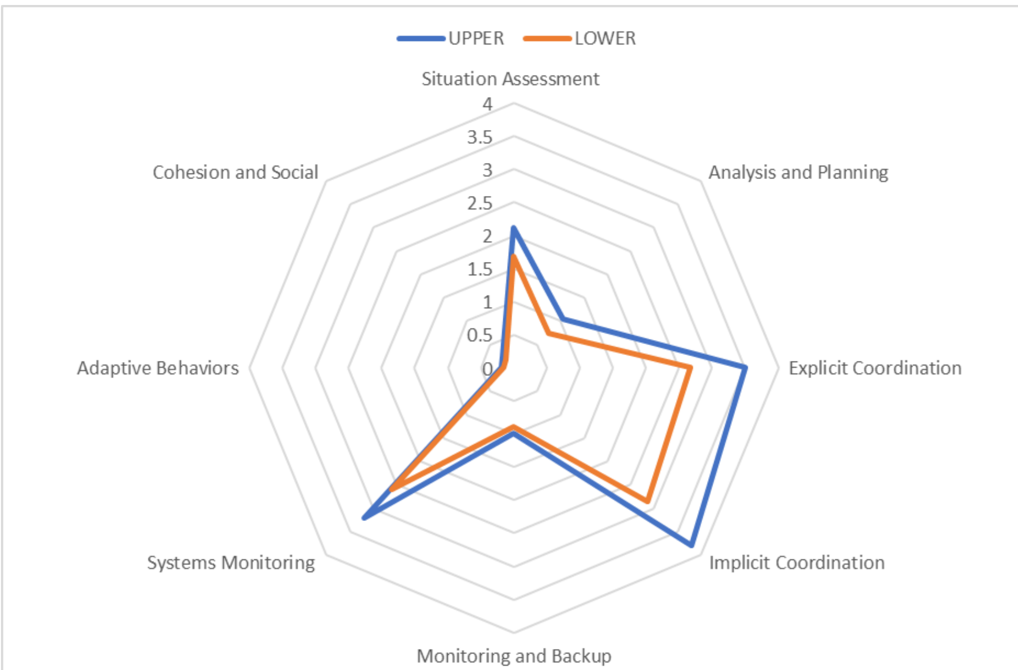


Figure 29. Frequency of teaming competencies between upper and lower performers in simulation genre

Table 67 presents test statistics and p-values for t-tests run to compare teaming competencies between upper and lower performers. Significant differences with large effect sizes are observed for explicit coordination (p = .046, Cohen's d= 1.2) and implicit coordination (p = 0.021, Cohen's d = 1.18). Significant difference with moderate effect size is observed for situation assessment (p = .042, d=0.61).

Table 67. t-test results comparing every teaming competency between upper and lower performers in Simulation. P values less than 0.05 are indicated with one star (*).

Teaming Competency	t-test (p-value)	Cohen's d
Situation Assessment	t (35) = 2.12, *p* = .042*	0.61
Analysis and Planning	t (35) = 1.32, *p* = .195	
Explicit Coordination	t (35) = 2.103, *p* = .046*	1.2
Implicit Coordination	t (35) = 2.4, *p* = 0.021*	1.18
Monitoring and Backup	t (35) = 1.43, p = 0.16	
Systems Monitoring	t (35) = 0.54, p = 0.54	
Adaptive Behaviors	t (35) = 0.74, p = 0.45	
Cohesion and Social	t (25.99) = 0.99, *p*=0.32	

RQ3 Discussion: Differences in Teamwork Behaviors Between Upper and Lower Performers

Differences in teamwork behaviors' frequencies were investigated between upper and lower performers. Several competencies were observed to be significantly different in more than one genre: situation assessment and coordination. The following discussion reflects on these differences, by exploring the implications of higher performers having a higher teamwork frequencies.

Situation Assessment Between Upper and Lower Performers

Frequency of teamwork behaviors was compared between upper and lower performers within every genre. Situation assessment frequency was significantly different between upper and lower performers within all three genres of puzzle platformer, asymmetric and simulation. Survival genre was not included in this analysis due to unclear performance criteria to separate upper and lower performers. Since time was one of the criteria to separate performance outcomes, this finding does not necessarily imply that upper performers had a higher count of situation assessment behaviors, but rather they engaged in more per minute. This can reflect the faster pace of higher performing teams in detecting cues and deriving their meaning. On a skill level, this can be tied to the individual's gaming expertise since individuals with higher expertise were observed to detect and understand environmental cues faster than novice individuals (Endsley, 2006). Additionally, since cue meaning is the behavior of understanding how the cue can affect the team's mission, a higher frequency of situation assessment can imply that higher performers engaged in meaning ascription more proactively and accurately, instead of taking more time to understand the cue.

Coordination Between Upper and Lower Performers

Implicit coordination was significantly different between the two performance levels within platformer and simulation genre. Previous work suggests that experts can rely on implicit coordination, since they have a shared team situation model, and therefore can rely less on explicit coordination (Ramo´ & Rico, 2008; Wuertz et al., 2018). In platformer and simulation, teams engaged in a variety of timing related tasks, where they had to manage interdependencies. Examples of implicit coordination include assembling a dish in Overcooked2 without verbally sequencing the steps or crossing a platform in BiPed without verbally exchanging which robot to go first. Teams who had a higher frequency of implicit coordination reflect a higher common understanding of the task. Furthermore, in simulation genre, explicit coordination was also significantly higher for the upper performers. This at first sounds counter-intuitive since higher performers are often found to rely less on verbal communication. However, simulation genres have limited opportunities for mutual performance monitoring since monitoring requires individuals to maintain an awareness of team-mates' progress. In simulation, this is hard due to the task allocation, and individuals being occupied with their own portion of the interdependent task. Therefore, a higher coordination frequency might suggest a higher frequency of information pushing. Pushing information is when team members provide information to teammates to support maintaining a common awareness of the task progress, while pulling is when team members must constantly ask about teammates' status (Demir et al., 2017; Hussain et al., 2008).

Examples of pushing information in simulation include reporting when observing an incoming enemy in Lovers in a Dangerous Spacetime or reporting task status in Overcooked 2 and Catastranauts (e.g., “I fixed the hole in the ship”; “rice is ready”; “there is a fire”).

Other Teamwork Competencies

No significant differences were observed in other competencies. This might have several implications. First, potentially the sample does not have adequate representation of teams with a large difference in expertise. It can be assumed that most of the teams have a moderate to high level of gaming experience, since the gameplay footage was posted by gamers. Additionally, it was observed through the gameplay that team-mates had a moderate to high familiarity with each other most of the time, and therefore they were not strangers playing as a team for the first time. Hence, potentially, the majority of analyzed teams had high levels of input, both on an individual and team level. Another implication indicates that frequency of behaviors is not necessarily the only indicator of the teamwork processes, but rather quality (Marlow et al., 2018; Marks et al., 2000; Manser et al. 2013). For examples, teams can engage in the annotated behaviors frequently, but that does not give a full picture on the quality of their teamwork, such as the accuracy of the information they are providing, or the quality of analysis and planning they are doing, or their proactivity in monitoring the environment in the right places. Therefore, future work incorporating a behavioral marker system can further assess the quality of the behaviors, to gain more insights on the teaming quality.

CHAPTER 7. SYNTHESIZED DISCUSSION: VALIDITY OF COOPERATIVE GAMES FOR TEAMWORK ASSESSMENT

This chapter synthesize the findings by exploring the internal validity of cooperative games for teamwork assessment. In this context, the validity of cooperative games aim to assess the suitability of using these environments as teamwork assessment testbeds, through exploring system validation. This study addressed three research questions to investigate consistency of cooperative games in inducing teamwork behaviors, associations between features and behaviors, and differences between upper and lower performers. The results found that cooperative genres induce homogenous teaming competency distributions for most competencies, while different cooperative genres induce different distributions. Therefore, cooperative genres affect teamwork behaviors, and can be used as a testbed design approach to target the desired teamwork distribution. Furthermore, cooperative features had different distributions of associated competencies, with some more frequently associated than others. Hence, the features are suggested as design guidance to emphasize the most frequently associated behaviors. Finally, the findings support the behavioral differences between upper and lower performers for several competencies, suggesting that teamwork is affecting performance outcomes in the studied games.

Synthesis of Research Questions: Cooperative Game Validity for Teamwork Studies

The use of cooperative games for teamwork assessment may be limited by validity (Mayer, 2018) and associations (Marlow et al., 2016). Related to system evaluation, validity is defined as “a confirmation through objective evidence that the requirements for a specific intended use or application of a system have been fulfilled” (Wilson et al., 2016, p. 03) . Therefore, if designers and researchers aim to use cooperative gaming environments to conduct teaming studies (e.g., measuring, eliciting, and training teamwork), the validity of these gaming environments must be assessed. We propose that the comparisons conducted within and between genres support several validity aspects. Jentsch & Bowers (1998) proposed an evaluation of the validity of pc-based simulations in studying aircrew coordination through demonstrating content and construct validity. Their evaluation approach was adapted here to fit the cooperative gaming purpose.

Content Validity

Content validity aims to answer whether cooperative games appropriately simulate activities that demand teaming behaviors. The underlying assumption of the study supported the claim that cooperative games are suitable for teamwork measurement, since they incorporate common goals, interdependent rules and mechanics, and communication support. The empirical results support this claim, by showing that video games across genres induced teamwork behaviors from all studied competencies. Therefore, the content of the video games replicated the task complexity and dynamic nature of teamwork environments, therefore inducing a variety of teamwork competencies, with different distributions depending on the cooperative genre (RQ1).

Construct Validity

Construct validity aims to answer whether the teamwork behaviors were related to the cooperative features in the games. Therefore, prerequisite for construct validity is satisfied if changing the gaming situations causes the teaming behaviors to also change. This validity was supported by the significant differences of behaviors between genres, and the significantly lower similarity measures of video games between genres explored in RQ1, which investigated the consistency of teamwork behaviors within and between genres. Additionally, associations presented in the area graph and bubble charts establish the most frequent associations between features and teaming behaviors in RQ2. Different cooperative features induced different distributions of teaming behaviors, therefore indicating that the features are influencing how teams allocate their teaming behaviors and triggering a range of varied behaviors. Hence, by changing the cooperative features, designers and researchers can affect the distribution and frequency of teaming competencies.

Another aspect of construct validity is convergent validity, aiming to answer whether designers can elicit behaviors related to their construct of interest. This was also supported, since by showing similarities within genres and differences between genres (RQ1), designers can elicit the intended behaviors based on their construct of interest by manipulating their cooperative design decisions.

Finally, construct validity also assesses whether the teaming behaviors were related to performance outcomes (RQ3). In other words, teamwork was necessary to perform well in these games. Since there was no direct access to teams' expertise levels, and there was a lack of low performing teams, this validity cannot be fully assessed. However, the performance-based comparison shows that several teaming competencies were significantly higher for upper performing teams compared to lower performing teams. Since there were not necessarily low performance teams, this analysis provides insights on the more frequent behaviors for the teams with higher performance outcomes, such as situation assessment, implicit coordination and adaptive behaviors being significantly different for puzzle platformer teams, with a moderate effect size, situation assessment for asymmetric teams with a large effect size, and situation assessment, explicit and implicit coordination for simulation teams. Therefore, teams with different performance outcomes exhibited different frequencies of certain behaviors, supporting that the teamwork behaviors were influencing their in-game performance.

Summary

In conclusion, the analyzed results provide support for both content and construct validity of cooperative video games for teamwork studies. Through exploring homogeneity, similarity, associations and performance differences, the results support the use of cooperative video games for teamwork assessment, and the ability to target teamwork competencies through design manipulations. The next chapter details design insights for every teamwork competency and concludes by synthesizing design frameworks for every cooperative genre.

CHAPTER 8. TESTBED DESIGN INSIGHTS

This section synthesizes the results, and builds on the discussion, to explore teaming competencies in further detail. For every teaming competency, the chapter reflects on differences between genres, and the underlying cooperative features driving these differences. Furthermore, it elaborates on how the findings can be tied back to teamwork research and attempts to explain why these cooperative features are inducing teamwork behaviors. Additionally, the sections build on the associations' results, to highlight the most frequent associations, and provide design insights on how cooperative features can be used to trigger teamwork behaviors. Finally, the chapter is concluded by presenting synthesized cooperative genre frameworks, that summarize how the cooperative features were observed within every genre in this study and concludes the takeaways of using these features to achieve the targeted teaming profile.

Testbed Design Insights for Teamwork Competencies

The following sections detail the design insights for every teamwork competency. The design insights aim to provide testbed design guidance, through understanding the factors that influenced the teamwork competency in the studied genres, based on the empirical teamwork profiles and associations.

Situation Assessment

Situation assessment percentages differed between genres, with puzzle platformer and asymmetric scoring an average of 23% and simulation and survival 13% and 14%. Therefore, there was a higher emphasis on situation assessment for puzzle solving based games (platformer and asymmetric), compared to the other two genres.

Situation assessment in video games

Situation assessment starts as an individual level behavior, where team members scan the surroundings to capture mission relevant cues (Endsley, 1995), and moves to forming a shared situation understanding, by deriving a meaning. In cooperative video games, players need to form a shared understanding of the situation, since they are working on interdependent tasks toward a common goal (Fullerton, 2014; Zagal et al., 2006). To understand the context of the mission, games implement graphical, auditory, or textual cues (Oulasvirta, 2009) in the game environment. In the analyzed cooperative video games, cues were considered any provided signal, symbol, mark, or object that supplies players with information about their environment and mission (Wuertz et al., 2018). Hence, it was first coded as an individual level process, where players are scanning for cues and recognizing them, and then generate a meaning and communicate it (Marks et al., 2001; Rosen et al., 2011).

Design insight: using puzzles to induce situation assessment through cue novelty

In puzzle platformers and asymmetric games, players were required to solve interdependent puzzles. As they navigate the game, every puzzle exposes them to cues relevant to their new challenge. In both genres, the top three features contributing to situation assessment were complementary and shared puzzles, and asymmetric obstacles and puzzles. For example, recognizing a yellow mark in It Takes Two, where one player needs to throw a nail for the other player to swing with their hammer and cross the platform (SA-CR: recognizing the yellow mark; SA-CM: understanding its meaning). Or recognizing a button in Portal 2 that should be reached to open the room’s door and clear the level. Therefore, by noticing a cue and communicating its meaning, teams assess the situation and form a shared understanding of the context. Cue meaning contributes to forming a team situation model, where team members form a common knowledge as they engage in their task (Richo et al., 2008; Cooke et al., 2000, 2003). Moreover, by exploring the associations between features and behaviors, situation assessment was repeatedly in the top competencies for puzzles, common risks, boss fight challenges and game pressures for both platformer and asymmetric, as these features require players to detect and understand cues to be able to execute the proper actions. In terms of frequencies, asymmetric genre was associated with more counts of situation assessment per minute, compared to puzzle platformer. This can be attributed to the pacing of the games. In puzzle platformers, players navigate a 3D environment, moving from one puzzle to another. When clearing a puzzle, players engage in platforming (running, jumping, swinging). Therefore, players are not only tactically engaged but also executively (they need to execute actions with their game characters) (Calleja, 2007) . While in the analyzed asymmetric games, teams were moving from one puzzle to another, mostly focusing on solving the puzzle as they go, putting more tactical efforts into the puzzle, and less executive efforts.

Design insight: effects of routineness of levels and decrease in cue novelty on situation assessment

As for simulation genre, teams engaged in situation assessment for 13% of their total behaviors, compared to 23% in puzzle platformer and asymmetric. The simulation game genres are characterized by a repetitive game loop (Sicart, 2015). And while all genres of games have game loops, the games categorized under simulation have one core game loop in repeat, for example players prepare and cook ingredients to deliver orders in Overcooked (this same gameplay is repeated throughout all levels), or players collecting resources and building train rails in Unrailed. This game loop is repeated in all levels with potential slight alterations, such as changing recipes, kitchen layouts and pressures in the game. Therefore, players do not need to constantly allocate their teamwork behaviors to detect cues in the environment, since the core cues become familiar due to the repetitive game loop (e.g., once a team understands that rice can be burnt if forgotten on the stove in Overcooked 2! they do not need to engage in cue recognition and cue meaning behaviors again). Situation assessment was rarely observed in the top competencies associated with simulation cooperative features, where it was only observed in top three for shared obstacles and awareness cues, emphasizing that it is not a top competency within the genre. Therefore, the lower percentage of behaviors dedicated to situation assessment in simulation genre compared to puzzle platformer and asymmetric can be attributed to the drop in the novelty of the cues and the increased familiarity of the situation due to the games' core game loop. When comparing frequencies, simulation games had a higher count per minute compared to platformer. While teams dedicate less of their behaviors (in percentages) to situation assessment, due to the fast-paced nature of the games, they engage in more behaviors per minute, as they are exposed to more frequent challenges, within time limits. The task allocation continuous process (TACP), through which the repetitive loop is implemented (e.g., delivering letters repeatedly, navigating space continuously), induced 85% of situation assessment. These levels happen fast in Overcooked, Unrailed, Key We and Cat astronauts, where players have 2-5 minutes to solve the level before moving to the next one. Therefore, the rate of exposure to new situations is higher. Additionally, team sizes in simulation games ranged between two and four, with three to four being the more dominant team size, while in puzzle platformer and asymmetric games, team sizes ranged between two and three, with two being the dominant size.

Design insight: the effect of a survival-oriented challenge system on situation assessment

Finally, situation assessment constituted an average of 13% of survival games' total behaviors. Survival games also rely on several repetitive loops such as the crafting game loop (players collect resources and craft utilities throughout the gameplay) (Sicart, 2015). Additionally, the games do not revolve around overcoming a puzzle-based challenge (like in puzzle platformer, asymmetric and simulation), but rather require players to survive in an open world environment, as they collect resources, overcome threats and craft tools and utilities. Shared utilities, crafting abilities and environmental resources were the top three cooperative features inducing situation assessment in survival games, indicating that players need to assess novel cues that are relevant to their survival mission (e.g., how to craft, how to use a resource). Therefore, the challenge system itself (surviving the environment) differs than the challenge system in the other genres and does not require players to constantly scan for novel cues and understand their meanings. Situation assessment was rarely observed in the top competencies of survival cooperative features, only observed for limited life resources, shared utilities, and awareness cues.

Situation Assessment Summary

In conclusion, situation assessment varied between genres. Puzzle platformer and asymmetric games demanded higher percentages, due to the novelty of puzzle cues, and the need to understand their meaning to solve the puzzles. Therefore, complementary, shared, and asymmetric puzzles can be used to emphasize situation assessment. In contrast, simulation genre required less percentages of situation assessment, due to the routineness of the task and drop in cue novelty as the teams repeatedly engage in the same task. As for survival, it also had lower percentages compared to platformer and asymmetric, due to the survival-oriented challenge system, and therefore the gameplay does not continuously emphasize on noticing new environmental cues and deriving their meaning.

Analysis and Planning

Analysis and planning competencies’ percentages were significantly different between genres, with survival requiring the highest percentage of 20%, and simulation the lowest percentage of 7%.

Mission analysis involves the team's collective discussions to interpret and evaluate the team's purpose. Teams aim to understand the goal of their mission and the main tasks that will contribute to its accomplishment, by incorporating the environmental conditions and available team resources (Marks et al., 2001). Planning involves establishing a course of action to complete the tasks.

Design insight: emergent interdependence encourages analysis and planning

In survival games, players engage in mission analysis to decide where they want to locate their base camp (e.g., discussing whether to locate it next to valuable resources or far from enemies), considering environmental conditions (resources, risks, and enemies), and available resources (utilities, weapons, and inventory). Teams can also develop plans to achieve the tasks of building a base camp (e.g., agreeing to build a base camp next to resources, generating a course of action to collect resources from the environment and bring them back to the decided location). Analysis and planning were repeatedly observed in the top competencies, having top spots for crafting abilities, shared utilities, environmental resources, common risks, community survival and team spirit, indicating the emphasis on collaboratively discussing the teams' strategies. Due to the emergent nature of the survival games, teams have alternate strategies to follow, therefore creating more room for analysis and discussions, and to develop plans. Specifically, survival games encourage experimentation, since players do not have a scripted progressive gameplay to follow (Bódi, 2021). In the stages of collaborative problem solving, teams engage in exploring and understanding, to interpret the initial information, then they represent and formulate through identifying approaches to solve the problem, and finally the plan through constructing a structure of actions (Graesser et al., 2018).

Design insight: interdependent puzzles engage collaborative problem solving

In puzzle platformer and asymmetric games, teams spent 11% and 12% of their behaviors engaging in this competency. In these genres, teams need to engage in collaborative problem solving to solve the puzzle. For example, in keep talking and nobody explodes (asymmetric), the expert (player with bomb diffusing manual) engages with the diffuser (player solving the bomb puzzles) to understand the mission. Since players have an asymmetric interface, they need to understand how to diffuse a puzzle, from the cues and meanings they generated. Analysis and planning were in the top competencies for asymmetric obstacles, puzzles, and environments. In Trine 4 (puzzle-platformer), players are in the same environment when they encounter a puzzle, and they would engage in mission analysis to figure out how to solve it. In Trine 4 players have complementary roles (e.g., sorcerer, warrior, thief), each with different abilities. Therefore, mission analysis was observed when players with different abilities were discussing how to approach the puzzle. Analysis and planning were observed in the top three for complementary puzzles and shared puzzles. Furthermore, teams can develop plans by laying out the next steps they will take, for example agreeing on grabbing a shield to reflect the light and hit the enemy in Trine 4 or develop a series of actions to open portals in Portal 2.

Design insight: fast paced games limit the ability to engage in analysis and planning

In simulation games, teams dedicated 7% of their behaviors under this competency. In simulation games, players are constantly in action phase. As soon as the game level starts, teams must start executing. For example, in Catastronauts, teams are in a spaceship that is constantly being shot at. Therefore, as soon as the level starts, team members need to start acting (e.g., shooting enemies, repairing ship cracks, grabbing fire extinguishers). In Overcooked2 teams have three to five minutes to deliver as many dishes as possible, they lose points if they deliver in the wrong order and food gets burnt if left on the stove unnoticed. Hence, teams dedicate less of their behaviors to engage in mission analysis, since they need to start executing immediately. Analysis and planning competency was rarely observed in the top competencies, only appearing for shared utilities and awareness cues, with low percentages of 6% and 9%. Since there is no down time in these games, players dedicate less of their behaviors to planning.

Analysis and planning summary

In conclusion, survival games had the highest percentage of behaviors in analysis and planning competency, and it was repeatedly observed in the top competencies of the genre. Since they induce emergent interdependence and don't have a linear game structure, players have the flexibility in deciding what to do next and how to do it. Players are provided with a variety of resources and a variety of possibilities to succeed with no imposed limits on how they should play the game. Analysis and planning were also emphasized in puzzle-oriented genres (platformer and asymmetric), appearing in top competencies for the puzzles, since they require collaborative problem solving. Finally, analysis and planning scored lower percentages for simulation games, due to the fast-paced chaotic nature, where less priority is given to analyzing, and developing plans.

Explicit Coordination

Explicit coordination occupied high percentages for all genres, ranging from 32% for asymmetric genre, and 15% for puzzle platformer. In explicit coordination, two behavioral markers were coded, each serving a different team purpose. First, explicit coordination-synchronizing and sequencing was coded whenever team members were engaged in interdependent tasks that required timing and synchronicity (e.g., the order of actions matter; team members needed to alternate their movements). While explicit coordination-reporting was coded whenever team members reported information on their status, actions, or environments, and this information contributed to the common task of the team (e.g., reporting what one player is observing in their environments in an asymmetric game; reporting what resources a player is finding in a survival game).

Design insight: Interdependent dynamics of timing and sequencing encourage explicit coordination

Explicit coordination reporting was observed to be triggered by the differences in individual experiences in the team. For example, in puzzle platformers, two players are in a shared environment, and they are exposed to the same information in the environment. Therefore, team members do not necessarily need to engage in reporting information, since the information is shared. However, there remains a need to sequence and synchronize actions in platformers, since the games have a variety of sequential tasks emerging from the closely interdependent tasks. For example, in BiPed, the two players cannot cross the color changing platform, unless they sequence their movements (one step from blue, followed by a step from red). Similarly in Portal 2, players are equipped with portals and cooperative communication mechanics (e.g., visual pointers), used to sequence or synchronize actions verbally or implicitly (Vaddi et al., 2015). Therefore, explicit coordination was repeatedly observed in the top behaviors of complementary and shared obstacles and puzzles (due to the sequencing and synchronizing focus), while it was observed in the top for common enemies, environmental resources, and shared environment (due to the need to report environmental updates, such an enemy incoming).

Design insight: environmental changes encourage reporting information to maintain a team situation model

In simulation games, even if players are in the same environment, there is a higher workload per person, where every person is engaged in a different task. Hence, team members engage in reporting their actions (e.g., rice is cooked; plate is delivered; I shot the enemies on the left side; I repaired the crack in the ship). By frequently communicating their actions and what they are observing, players contribute to forming a team's situation awareness, and reduce the cognitive workload on each other, by pushing information, instead of requiring players to pull information (e.g., asking about the status of the ship; asking about the status of the rice; asking for reports on the locations of enemies in Lovers in a Dangerous Spacetime). Additionally, these games incorporate a variety of closely interdependent sequential tasks, such as typing letters in a specific order in KeyWe. Explicit coordination was repeatedly observed in the top behaviors for almost all features in the simulation genre. Similarly, in asymmetric genre, players had to sequence or synchronize interdependent tasks with an asymmetry of information (e.g., expert player telling the diffuser in what order to press the buttons in a Simon says game in Keep talking; the hacker guiding the spy in navigating security drones in Tango). Therefore, as one player has access to information that the other player needs to execute their task, players engaged in sequencing and synchronizing actions, by providing information that directly affect the other player's actions, and ultimately the team's mission (Handler, 2017; Dormer et al., 2017). Additionally, players need to constantly report any environmental changes, such as how much time is left, associated with the timer (game pressure), which was observed to influence communication dynamics in keep talking (Fine, 2016).

Design insight: dispersed environmental resources encourage reporting information

In survival games, since players are dispersed across the environment, they constantly report what resources they are finding (e.g., I found a field of gold in the south; I am collecting wood now; I built a science machine). For example, EC was in the top behaviors for environmental resources, common enemies, and shared environment, ranging from 23% to 29% of the behaviors induced by these features. Since the players are dispersed across the environment, they need to repeatedly report their status updates.

Explicit coordination summary

In conclusion, all genres demanded high percentages of explicit coordination, with emphasis on either reporting, sequencing, and synchronizing, or both behavioral markers. In simulation, asymmetric and platformer, the interdependent puzzles required sequencing and synchronizing of different levels. It can be required sequentially (player one complete action before player two starts action), coincidentally (players one and two must perform at the same time, such as pressing two buttons), concurrently (players continuously perform together, such as navigating a ship with one steering and one shooting enemies in It Takes Two; or players concurrently washing dishes and cooking rice in Overcooked), or expectantly (player one can start their move if player two is ready and standing by, such as having to alternate movement in color changing platforms in Biped, and red robot cannot move unless blue robot is ready to move next). Other aspects of the games required explicit coordination-reporting, such as survival games with a need to update status due to the dispersed environment, or simulation due to the constant changes in team-mates status and the overload, or asymmetric due to the asymmetry in information.

Implicit Coordination

Implicit coordination is when teams engage in coordinating their interdependent tasks, without verbally speaking. The competency included two behaviors, one for implicitly sequencing and synchronizing actions (IC-S) and one for anticipating, by doing actions that contribute to the team's advancement without verbal demands (IC-A). Implicit coordination had the highest percentages in simulation genre (26%), and lowest percentages in asymmetric (2%).

Design insight: routineness in core tasks provide opportunities for implicit coordination

Previous work suggests that under high workload conditions, teams rely more on implicit communication instead of explicit communication (Salas et al., 2007). In simulation genre, in addition to the high workload, teams are exposed to familiar situations repeatedly, as described previously in the concept of the repeated core game loop. This repetitiveness can be linked to task routineness, defined as tasks with high routines that are well defined, structured and with predictable situations (Rico et al., 2008). Therefore, teams dedicate 26% of their behaviors to implicit coordination, which is the highest percentage of behaviors in simulation genre. For example, once teams understand how they are assembling the sushi dish in Overcooked 2 they start engaging in implicit coordination, where players put nori, and then rice, and fish and deliver to the customer, without needing to communicate. Same in Unrailed, where once players understand how they are removing obstacles and collecting resources for the train tracks, they start cutting trees, mining bricks, building, and assembling trails without communicating. As task routineness increase, there is more room for teams to be able to efficiently engage in implicit coordination. However, in the task allocation continuous feature, players engaged in implicit coordination as equally as explicit coordination. While simulation genre provide room for high routineness through the repetitive task structure, the game environment is continuously changing with uncertainty aspects, such as taking ship damage in Catastronauts, encountering unexpected enemies in Lovers, or not knowing what the next recipe will be in Overcooked 2, which also emphasizes the need for explicit coordination. Another aspect of implicit coordination in simulation genre is the anticipation behavior. Individuals are more prompted to act to help the team advance, without the need to be told to do so, since the situation is an emergency. For example, implicit coordination constituted 62% of the behaviors associated with shared utilities in this genre, since players will use them in anticipation of the next step (e.g., preparing trails in advance to avoid crashing the train, preparing, and getting ingredients from the inventory, fixing a crack, or setting off a fire in the ship).

Design insight: familiar design patterns across levels allow for implicit coordination

Puzzle platformer also induced 20% implicit coordination in its competencies. As described in the explicit coordination section, these games involve interdependent puzzles with timing and synchronicity demands, and while every puzzle exposes players to new cues, the core mechanics can be similar, such as using the hammer-nail combination to swing across platforms repeatedly in It Takes Two, or opening portals and crossing in Portal 2, or shifting sizes to lift each other in Shift Happens, therefore allowing room for implicit coordination. IC was repeatedly observed in the top competencies of the cooperative features, occupying 30% for complementary obstacles and 28% for shared obstacles.

Design insight: asymmetry provides low opportunities for implicit coordination

In contrast, asymmetric video games had an average of 2% implicit coordination. This is an expected result. Since these video games entirely rely on asymmetry, there is minimal room for players to anticipate the next action that will benefit the team or sequence their actions implicitly.

Monitoring and Backup

Monitoring and backup had a range of percentages between genres, scoring the highest average of 13% in survival, and lowest of 3% in asymmetric. This competency involved the behaviors of mutual performance monitoring, which involve individuals observing other players directly to monitor their actions and task progress, or verbally asking on individuals' status. In addition, the behaviors of backup behaviors (including feedback, resources, and behavioral help), whether proactive (without being asked to), or reactive (answering a teammate's request).

Design insight: workload diversity and limited life resources encourage monitoring and backup

In survival games, monitoring and backup was observed in the top competencies for common enemies, limited life resources, sharing abilities, team awareness cues, and team spirit, scoring 48% of the behaviors associated with the latter. In this genre, there is no mandatory need for cooperation, therefore, a variety of the player's behaviors to ensure teammates survival emerge from the dynamic labeled team spirit, which indicates that the behavior was not triggered mandatory, but rather from the player's desire to ensure the team's survival. For example, individuals provided help to teammates when enemies attacked them, or when they were starving and needed food or revival. Hence, survival environments created more need for backup behavior, since there is no fixed workload assigned to the players by the game, but rather individual's status defer depending on their individual gaming experience (e.g., starving, being attacked by an enemy). Therefore, this imbalance in experiences, trigger opportunities to provide backup.

Design insight: tightly coupled interdependent tasks allow for monitoring through direct observation

As for puzzle platformer, monitoring and backup was primarily induced by the complementary and shared puzzles, and it was occasionally observed in the top competencies induced by the cooperative features, only appearing in the top three for complementary puzzles and common risks and occupying 100% of the behaviors induced by limited life resources. In this genre, since the tasks are highly interdependent, and one player's progress strictly depends on the other's progress, members can be inclined to monitor each other. Monitoring facilitates maintaining an awareness of one another's activity and moderate the synchronization of actions (Robert, 2016) . Particularly, if there is a difference in skills, team members might engage in more monitoring to ensure that other teammates are performing their actions as expected, which is inversely related to cognitive trust, a trust developed between team members based on perceived competency and ability (Robert, 2016). However, in the annotated platformer footage, differences in skills levels were not frequently observed among team members, therefore potentially limiting the need for monitoring and backup.

Design insight: high workload with dispersed task stations lowers the ability to monitor team members

Finally, monitoring and backup scored relatively low in simulation genre with an average of 7%. As described in previous competencies, while the tasks are interdependent, individuals are occupied with doing their parts of the tasks, lowering their priority and ability to monitor each other and provide backup, especially that monitoring requires continued attention and contact between team members (Dennis et al., 2012), which is difficult to maintain in high workload environments like the analyzed simulation games. However, MB was observed in the top competencies for sharing abilities and limited life resources. For example, when individuals are lacking resources to execute their tasks, other individuals can share resources with them, or when character's life resources get low in Catastronauts, other individuals can notice and advise them to replenish. Hence, while simulation games involve incidents of different workloads, therefore theoretically providing chances for monitoring and backup, due to the excessive workload on all players, it is less likely for players to be capable of constantly monitoring and helping each other.

Design insight: features that allow direct monitoring in asymmetric environments can encourage monitoring and backup compared to total asymmetry

As for asymmetric games, there is no ability for individuals to provide resources or behavioral backup (such as directly taking someone's tasks), and potentially little opportunities to provide feedback backup (since individuals do not know how to solve the other individual's obstacles). However, even verbal monitoring was rarely observed, such as asking other players for their status. This can be explained by the similar levels of expertise, and the individuals' proactivity in reporting (discussed in the high percentages of explicit coordination in this genre), therefore pushing information on their status, instead of requiring the other players to pull information by asking them about their status. However, monitoring and backup was observed in the top behaviors associated with team awareness cues. This is a feature that allows players to directly monitor each other even in an asymmetric environment, which was observed in Operation Tango, where the hacker can occasionally access the building's cameras to monitor the spy's location and provide help. This is also supported by the higher percentages and frequencies of monitoring and backup in Operation Tango compared to other asymmetric games.

Systems Monitoring

Systems monitoring had similar percentages for simulation and survival (21%) and asymmetric (18%), while scoring lower in puzzle platformer (11%). This competency involved two behaviors, one for external systems monitoring (monitoring all elements hosted in the external environment hosting the team), and internal systems monitoring (monitoring the internal team generated resources). In cooperative video games, team members need to develop and maintain a Game-space awareness (GA), where they need to be aware of the conditions of the game environment, including items and resources (external), and what changes are happening to the environment due to team member's actions (internal) (Teruel et al., 2016). Hence, systems monitoring is closely related to developing and maintaining GA, and different genres provide different environments to do so.

Design insight: constant environmental changes encourage internal and external systems monitoring

For simulation genre, systems monitoring was frequently observed in the top competencies induced by the genre's cooperative features, including task allocation continuous process, common risks, game pressures, dynamically changing environment, and environmental resources. While there is less emphasis on monitoring and backup due to the overload of individuals, there is a higher need to constantly monitor the environment since it is continuously changing. For example, in Catastronauts, the ship gets damaged at a very fast pace, therefore teams need to keep track of cracks, fires and damage. In Unrailed, the train is constantly moving, and team members need to monitor for any obstacles and the way, and any depleted resources that need replenishment (Grandi, 2021). In Lovers in a Dangerous Spacetime, there is a particular need to monitor the external environment, since the ship is continuously attacked by enemies, and encounter space meteors and risks. Similarly, the competency frequently appeared in the top competencies in survival genre, including common enemies, environmental resources, common risks, and shared environment.

Design insight: dispersed environmental resources encourage systems monitoring

While in Simulation the motivator is the constantly changing resources, risks, and environmental status the individuals are exposed to, in survival genre it is the static yet dispersed resources and risks that trigger system monitoring. Particularly, Don’t Starve Together scored the highest frequency of systems monitoring within survival games, with an average of 1.58/minute, and the highest frequency in all other teaming competencies compared to survival games in the genre. This is relative to Don’t Stave Together’s more pressured rhythm compared to the other survival games, where it has shorter day/night cycles (Costello, 2018). For example, individuals need to monitor the environment by looking around to find resources, or to avoid enemies, and track day status (e.g., in Don’t Starve characters go insane in the dark), or to go and mine resources in Atroneer. Teams can track three types of information in gaming environments: procedural, episodic, and factual (Belanich et al., 2004). Observing the game environment allows team members to gather episodic information.

Design Insight: asymmetry encourages verbal systems monitoring to monitor the inaccessible environment

As for asymmetric games, systems monitoring was happening through both behavioral actions (player characters monitoring their own environment), and verbally (players monitoring the other environment through asking questions). Systems monitoring was repeatedly observed with asymmetric obstacles, asymmetric puzzles, common risks, game pressures, and asymmetric environment, scoring 46% of the total behaviors associated with the latter.

Finally, for puzzle platformer, less emphasis was placed on systems monitoring in the dominating features (such as puzzles and obstacles), however it did score high percentages in association with common enemies, environmental resources, shared environment, dynamically changing environment, and boss fights, ranging from 42% to 53% of the total competencies associated with these features. Hence, a repeated pattern can be observed with systems monitoring across genres, where it was observed to be frequently associated with environmental elements, resources, risks, and pressures, across all genres.

Adaptive Behaviors

Adaptive behaviors scored low percentages across all genres, ranging from 6% in asymmetric to 1% in simulation genres. However, in terms of frequency, it did score up to 0.19/minute in some simulation games (equivalent to approximately one competency every five minutes) and ranging from 0.41 to 0.51/minute in asymmetric (approximately one competency every two or three minutes). Therefore, the behaviors were happening throughout the gameplay, however at a slow rate, and occasionally. This can be associated with several explanations.

Design insight: reviewing while performing is limited due to the fast-paced nature of video games

First, the coded adaptive behaviors included reactive strategy adjustment, and team learning-reviewing. As for the latter, this behavior potentially is less likely to happen during the gameplay, but rather before, between and after the sessions. This was observed in some video games, where teams would review their performance while moving from one level to another in some simulation and asymmetric games, therefore explaining the slow pace (it occurs between levels). Another reason can be associated with the fast pace of games in several genres, particularly simulation, where there is little to no room to review performance while executing. As for reactively adjusting strategies, the code was aimed to capture when teams decide on a course of action, and then verbally change the course of action by deciding on a new plan.

As for the underlying features associated with competency, puzzles induced most of the adaptive behaviors in both platformer and asymmetric genres. Similarly, the continuous process in simulation induced 89% of the adaptive behaviors. Hence, when it occurred, it was induced by the main tasks of these three genres (the puzzles or the continuous loop), and it did not appear in the top competencies of any of the features. Therefore, potentially the competency was masked by other competencies, due to its limited need (reviewing performance and developing new strategies).

Cohesion and Social

Cohesion and social behaviors ranged from 2 to 4% across genres, also scoring relatively low percentages. This competency coded team cohesion (encouragement and complimenting) and social interactions (conversations). However, they were only coded when they were triggered by a game feature, and they were coded on a team level (e.g., great teamwork, we are experts, great job). By looking at the frequencies, the competency scored up to 0.42/minute in Overcooked 2! (Approximately one every two or three minutes), and 0.55/minute in keep talking, and 0.4/minute and 0.44/minute in it takes two and portal 2. Therefore, the behavior was happening across the gameplay for several games, however, was dominated by the most frequent teaming behaviors, especially that the behavioral markers were not designed to capture granular social interactions in this study, such as individuals calling each other's by names, or having conversations irrelevant to the game context, or verbal commentary interactions streamers might do to address their audience.

Design insight: story and aesthetics encourage social interactions

By looking at the features, cohesion and social was the top competency for aesthetics and story across all gaming genres (e.g., players commenting on the games' aesthetics, players having a conversation regarding the story). It also appeared as top competency in competitive challenges which appeared in It Takes Two for example (where players play little side challenges where they compete instead of cooperating). These challenges were observed to induce laughter and friendly competitiveness between players. However, by looking at the breakdown of the competency itself, 31% of its total was induced by complementary puzzles in platformer genre, 48% by asymmetric puzzles in asymmetric genre, and 95% by the continuous task in simulation. Therefore, this shows that cohesion and social was present across the gameplay, such as players expressing compliments and motivation toward each other when overcoming challenges, or celebrating when overcoming a challenge, even if with low overall percentages and frequencies when compared to other teaming behaviors.

Design Insights Summary

Teamwork competencies varied in distribution within cooperative genres and between cooperative features. Therefore, design insights derived from the genres and features allow for further guidance on how to use them as design features for teamwork assessment testbeds to emphasize the desired teamwork competencies. Table 68 summarizes the described design insights. The previous section reflected on the teamwork competencies' associations with the cooperative genres and features and suggested a variety of design insights derived from the studied features. This section summarizes the design insights through visual representations of the observed cooperative patterns, and elaboration on how to use the patterns to induce teamwork behaviors. The visual summaries are provided to facilitate referring to the design features and as general guidelines for testbed design.

Table 68. Summary of design insights associated with teamwork competencies

Teamwork Competency	Design Insights
Situation Assessment	1. Encouraged through *cue novelty* 2. Lower emphasis with *routineness of tasks* and *decrease in cue novelty* 3. *Survival oriented challenges* require less emphasis on cue recognition and meaning
Analysis and Planning	1. Encouraged through *emergent interdependence* 2. *Interdependent puzzle* solving requires analysis 3. *Fast paced games* lower the opportunities for analysis and planning
Explicit Coordination	1. Encouraged through *interdependent sequencing and timing* 2. Encouraged through *constant environmental changes* 3. Encouraged through *dispersed environmental resources*
Implicit Coordination	1. *Routineness in core task* supports implicit coordination 2. *Familiar game patterns* support implicit coordination 3. *Asymmetry* provides low opportunities
Monitoring and Backup	1. *Workload diversity* creates opportunities for backup 2. *Closely coupled interdependent* tasks allow for direct monitoring 3. *High workload with distributed tasks* limits monitoring abilities 4. *Monitoring features* in asymmetric games can create opportunities for backup
Systems monitoring	1. Encouraged through constant *external and internal environmental changes* 2. Encouraged through *dispersed environmental resources* 3. Verbal systems monitoring in *asymmetric features*
Adaptive Behaviors	1. *Fast paced games* limit learning and reviewing while performing
Cohesion and Social	2. Encouraged through *story and aesthetics*

Visual Summary of Cooperative Design Patterns

This section summarizes the design insights through providing visual representations of the observed design patterns and elaborating on how to use them to induce desired teamwork distributions and behaviors. Every visual summary is composed of a variety of visual icons, representing cooperative design features and patterns observed in the study, including players, player abilities, interdependence, environmental resources, threats, and core game loops. This synthesis is achieved through the cooperative feature tracked in this study.

Puzzle platformer

Puzzle Platformer video games were the most diverse genre in implementing cooperative features. While puzzles dominated the game play and were the most frequently associated with teaming behaviors (complementary obstacles and puzzles, shared obstacles, and puzzles), the games implement a variety of environmental components including common enemies, common risks, and boss fights. As discussed in the competency section, puzzles in puzzle platformer genre engage teams in new challenges, with new cues to assess and derive meaning, to be able to solve the puzzle and figure out how to proceed. Figure 30 represents a visual representation of the synthesized framework, highlighting the common design patterns in the analyzed puzzle platformer games. Teams navigate an environment moving from one platforming puzzle to another. These puzzles require players to engage in interdependent tasks with different types of timing and synchronicity (e.g., sequential, concurrent, coincident (Harris, 2019). The puzzle environment constitutes of components needed to solve the puzzles, which inherently provide cues to the players on how to solve the puzzles. The common goal within puzzles is to clear it to move to the next one. Within the gaming environment players can encounter resources, enemies, boss fights, and threats.

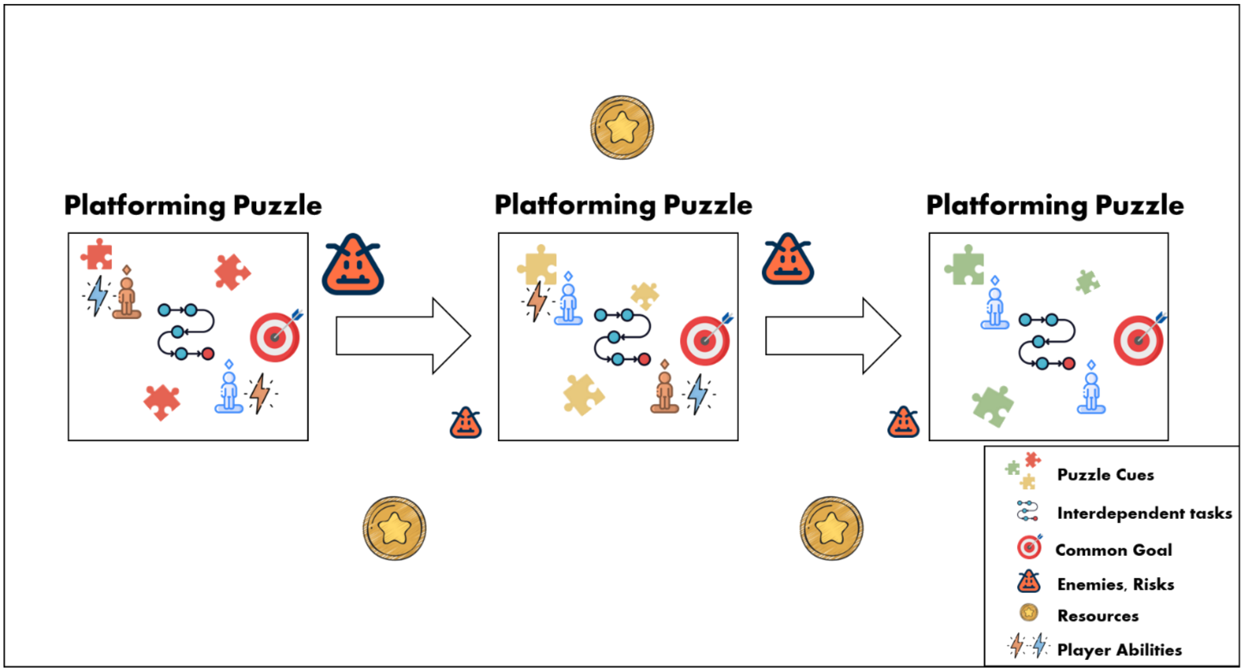


Figure 30. Visual summary of cooperative design patterns within puzzle platformer genre

Cue Novelty is one factor contributing to high percentages of situation assessment in puzzle platformer genre. It is represented in the Figure 30 through the puzzle cues with different shapes and colors. As for coordination, both explicit and implicit coordination had high percentages within the cooperative features of this genre. As previously discussed, *interdependent tasks with timing and synchronicity* require coordinated actions (whether explicit or implicit). It is illustrated in Figure 30 through the interdependent task's symbols. Other competencies, such as monitoring and backup, account for 13% of the pairs associated with complementary puzzles, and 12% of the pairs in shared puzzles. Due to a *shared environment and interdependent tasks*, players can monitor each other's behaviors and verbalize it by providing guidance or commenting on someone's performance. Furthermore, *complementarity* of abilities increased the interdependence, by forcing players to rely on each other. It is presented in Figure 33 through the player abilities signs with different colors. As for common enemies, environmental resources, shared environments and boss fights, the biggest bubbles were with systems monitoring, with percentages ranging 42% to 53%. Features that heavily rely on *shared environment components* (e.g., tracking resources in the environment, tracking enemies in the environment, tracking boss fight attacks in the environment), appear to have the biggest portion of their associated behaviors under systems monitoring. They are presented in Figure 30 through the Enemies, Risks and Resources symbols.

Asymmetric

The analyzed asymmetric genre included games that are *strictly asymmetric*, where players were separated by *environment, information, and roles*. The provided framework synthesizes the cooperative features observed in this genre in this study, where asymmetric puzzles and obstacles, that rely on interdependent tasks, dominated the gameplay. Figure 31 presents a visual synthesis of the observed cooperative features.

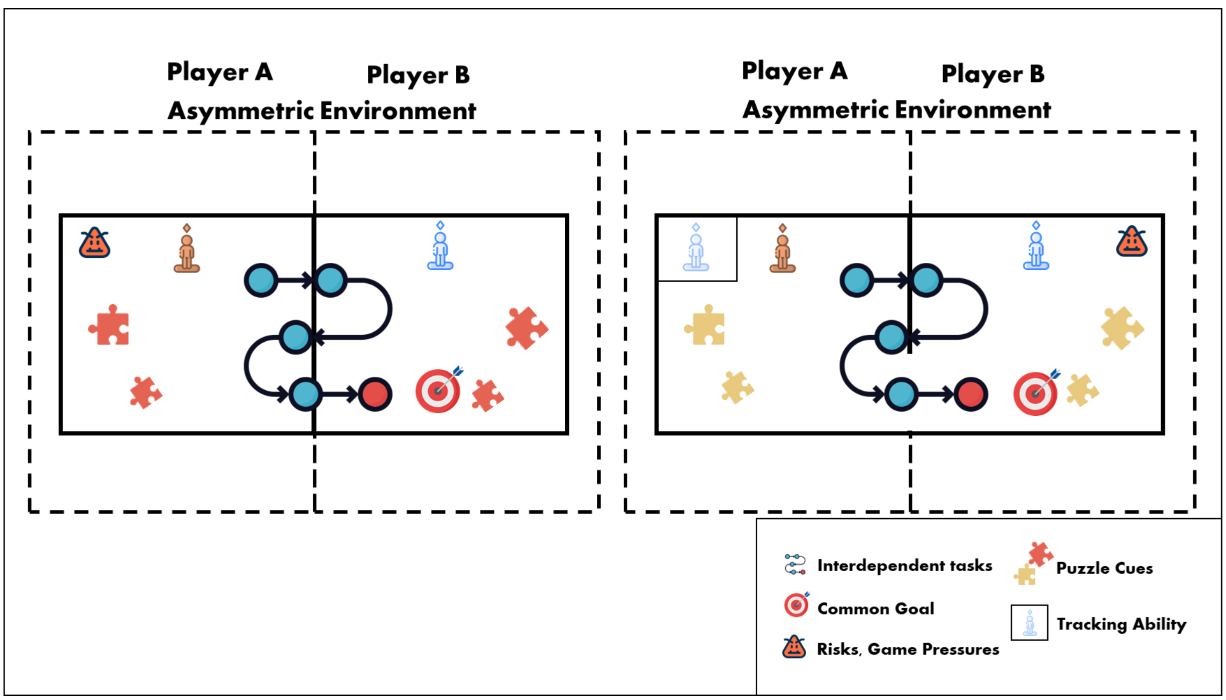


Figure 31. Visual summary of cooperative design patterns within asymmetric genre

The top competencies associated with asymmetric puzzles and obstacles included situation assessment (28-26%), explicit coordination (28-33%), and analysis and planning and systems monitoring (13%). Particularly, these puzzles require players to *verbalize* most of their teaming behaviors, including cue recognition and cue meaning, reporting, sequencing, and synchronizing, and verbal systems monitoring to keep track of the environment they do not have access to. The asymmetric environment is illustrated in Figure 31 through the complete visual separation of Player A and Player B. Some video games provided team *tracking abilities*, which allowed for performance monitoring, and visually monitoring the other environment (rather than verbally). This feature can be controlled by designers to investigate differences when team members are allowed to visually monitor each other in asymmetric environments or not. It is illustrated in Figure 31 through the Tracking Ability symbol. Furthermore, some video games (like we were here forever), allow players to *exchange environments* occasionally, indicated by the gap in the second puzzle in Figure 31. This is another feature that raises further exploration opportunities, by allowing individuals to occasionally exchange environments instead of absolute asymmetry. Additionally, the games have game pressures, such as timer in Keep Talking, and common risks (such as firewalls in Operation Tango), that can have effects on the game rhythm, as seen in Keep Talking where the frequency of behaviors is higher than the other asymmetric video games. Designers can use game pressures to manipulate the frequency of teamwork behaviors and increase urgency.

Simulation

Figure 32 visually summarizes the cooperative patterns observed in the Simulation genre.

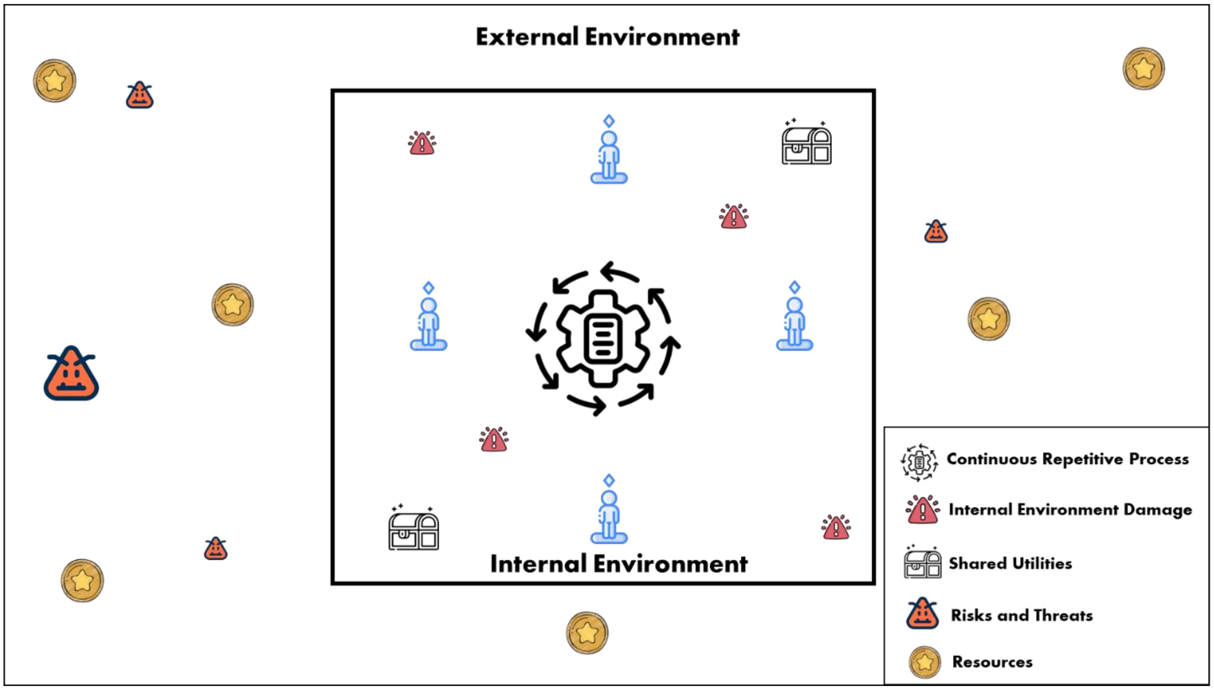


Figure 32. Visual summary of cooperative design patterns within simulation genre

Simulation video games were characterized by a *continuous repetitive process*, which increased the *routineness* of the task, and allowed for high percentages of *implicit coordination*. This is illustrated with the Continuous Repetitive Process icon placed in the middle of the visual in Figure 32. However, explicit coordination was still highly required due to the constant changes in the environment and potential *internal environmental damages* that can arise, such as incoming enemies, train overheating, ship damage, and burning foods. The visual icons are represented in Figure 32 to illustrate this pattern. Therefore, the feature balanced routineness of the core task with continuous changes of the context, to create opportunities for both implicit and explicit coordination. These dynamic changes in the environment also required high percentages of *internal and external systems monitoring*, complemented with *pushing information* to report environment status updates. External threats, risks and enemies were observed to drive external systems monitoring, when present, such as in Lovers in a Dangerous Spacetime and Cat astronauts. Alternatively, the continuously changing internal environment status was observed to demand the internal systems monitoring behaviors. Additionally, the teams were provided with shared utilities, such as food inventories for cooking, tools to repair the ship, new weapons to defend enemies, and train tacks for assembly.

Survival

The survival genre was driven by a communal need for survival, further encouraged through team spirit and community survival. These two features were observed to encourage teamwork behaviors even when interdependence was not mandatory. These features were characterized by providing players higher chances of survival if they ensure the team's and community survival. Figure 33 summarizes the observed design patterns in the survival genre.

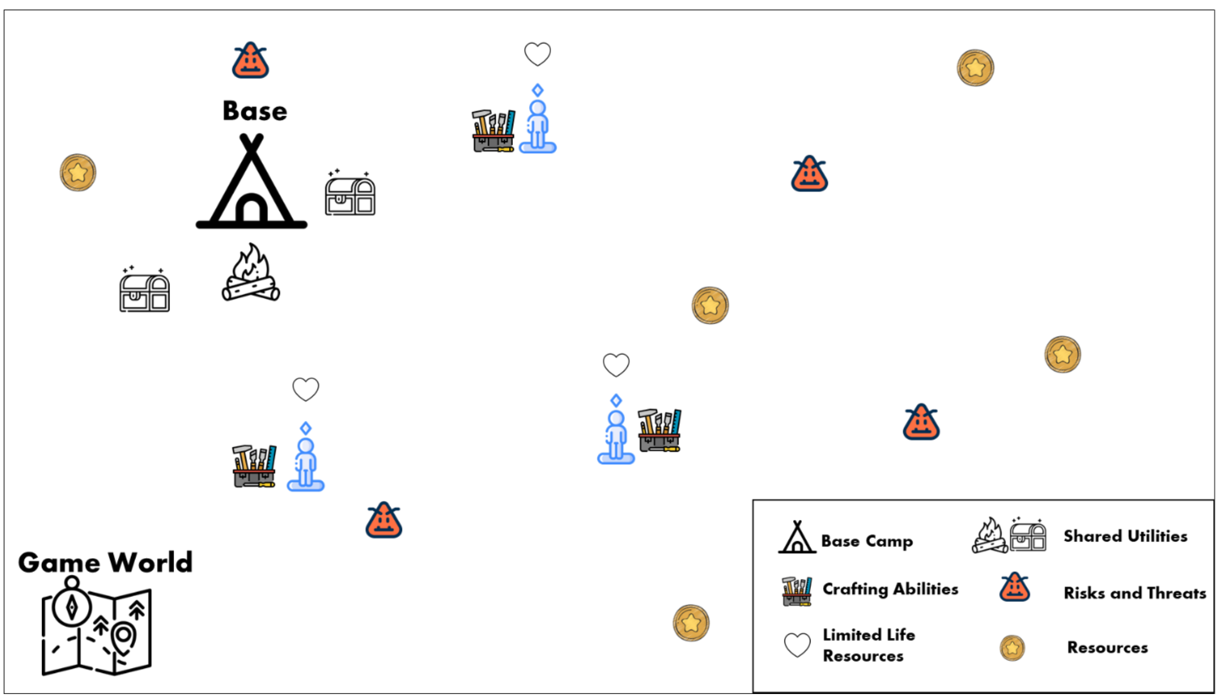


Figure 33. Visual summary of cooperative design patterns within survival genre

The survival genre provided players with *largely distributed environmental resources*, therefore players were required to *divide and conquer* to cover more territory and engage in external systems monitoring. This is illustrated in Figure 33 through the dispersed Resources. Furthermore, *common risks and threats* encouraged teamwork behaviors, such as everyone going insane when it is dark in Don't Starve Together. Additionally, the genre equips players with *crafting abilities*, providing opportunities to *craft shared utilities* and *build common base camp.* Furthermore, individuals had *limited life resources*, which derived backup behaviors, such as players sharing food when someone is starving.

Survival genre was the only genre where individuals were individuals mostly engaged in emergent interdependence while being loosely coupled. This emergent interdependence provided a variety of possibilities for players to execute their actions, such as discussing where to locate base camp, analyzing different locations based on resources, and elaborating on what could be the best approach to follow, therefore allowing high percentages of analysis and planning. Team members were provided with more opportunities to execute their own tasks, and some teams were even observed to do more individual strategies. However, most teams were observed to be oriented toward teamwork. In survival genre the framework seems to work in cohesion, rather than singular cooperative features driving behaviors. While iterating the codebook, the cooperative features were harder to separate in this genre, since potentially multiple features are *implicitly motivating players behaviors*, such as their continuously depleting life status, or the continuous perception of common threats. And hence, creating future research opportunities to explore how these cooperative features (e.g., crafting abilities, environmental resources, limited life resources, common risks) would induce behaviors when isolated from other features, or when combined.

CHAPTER 9. CONCLUSION

Summary

This work aimed to develop and employ an observation-based teamwork measurement system in cooperative video games. The process involved developing a system to track behavioral markers and cooperative features in cooperative gameplay footage through developing a codebook. The codebook development went through several phases of initial code sourcing, iterations to refine the codebook, inter-rater comparisons, and implementation. The video game selection process involved cooperative and technical criteria. The study analyzed four cooperative genres: puzzle platformer, asymmetric, simulation and survival. Each genre was categorized based on the major cooperative features driving the players interactions with the game. A total of 177 teams were analyzed across 18 games, categorized under the four genres. Through annotating teamwork behaviors in association with a cooperative feature, the study generated teamwork profiles, to assess similarity and differences within and between genres. Furthermore, counts of associations were analyzed to generate the most frequent associations and provide design insights to trigger teamwork behaviors through cooperative features in teaming testbeds. Finally, comparisons between upper and lower performers were conducted to gain insights on behavioral differences.

Contributions

The study had several research contributions. First a teamwork measurement system was developed through a codebook of teamwork behavioral markers and cooperative features for application in cooperative gamified environments. This codebook builds on existing teamwork models and adapts behavioral markers to be observable and trackable in cooperative games. The codebook allows researchers to capture dynamic teamwork processes in these environments and provide a wider understanding of teamwork profiles in these games. In future work, the codebook in conjunction with other measurement systems. By collecting data from multiple resources, researchers can triangulate to develop a richer understanding of teaming behaviors. A more robust teamwork measurement system can be developed by pairing the codebook with self-report measures.

Additionally, the study demonstrated similarity of teamwork competencies' distributions within cooperative genres, and different patterns of behaviors in different cooperative genres. Through this finding, designers can use cooperative genres to target the desired teamwork profile. The study provides teamwork profiles within every genre, indicating that specific cooperative genres can be used to trigger the desired teamwork distributions.

Furthermore, the codebook includes a list of cooperative features, derived from literature, and further expanded on through an inductive approach, therefore providing an approach to study associations between cooperative features and teamwork behaviors. The thesis demonstrates the annotation of these cooperative features and provides the associations between cooperative features. The cooperative features were further synthesized into design frameworks, where they work in harmony to induce teamwork behaviors. The explored associations and design frameworks provide development guidance for designers and researchers. Developing teamwork assessment testbeds, can benefit from implementing a variety of cooperative features, to study the targeted teamwork construct. The approach of triggering teamwork behaviors through design features is consistent with existing teamwork testbed design methods, such as the event-based approach to training (Fowlkes et al., 1998; Rosen et al., 2008).

This study contributes to a cooperative gaming approach on EBAT, where designers can use cooperative gaming features as critical events that can trigger the desired behavioral markers.

The approach of systematically implementing cooperative features to trigger teamwork behaviors can have several benefits. Benefits include providing a structured observation technique, a systematic introduction of cooperative features to trigger behaviors, and an accurate assessment of the desired constructs (Salas et al., 2017). Furthermore, the study clarifies the relationship between cooperative attributes and teamwork behaviors, a step necessary for the continued use of cooperative games for teamwork studies (Marlow et al., 2016) . Additionally, this future research opportunities, to explore how these cooperative features can work in isolation (if separated from their design framework), or in cohesion with other cooperative features from different genres.

Finally, the study explored the internal validity of cooperative video games, for teamwork assessment, through consistency, associations, and performance comparisons. By supporting validity, the study provides empirical support for using cooperative games as assessment environments and learns from their design features to develop teamwork testbeds. Assessment design relies on the environment's ability to induce observable behaviors that can be associated with the construct of interest. The study demonstrated this process through the consistency of profiles of the tracked behaviors, and the empirical associations.

Limitations and future work

This study had several limitations. First, observational teamwork measurement is susceptible to several drawbacks, including errors associated with human judgment, and potential bias and subjectivity that can arise with unfamiliar or vague behaviors (Roberts et al., 2022). Specifically, the final application of the codebook involved only one coder. While this can minimize confounding effects and inconsistencies that can arise from multiple coders, it can affect the completeness of the coding, since the results did not incorporate findings from multiple coders. Thus, some events may have been missed.

Repeatability of the coding process was addressed with the inter-rater agreement process used during the development and refinement of the codebook. However, the inter-rater agreement process is also subject to some limitations, including that the formula used accounted for agreements over total codes, which does not account for chance and missing data. Furthermore, the inter-rater coding was obtained based on sum of agreements in a section of time, rather than a granular comparison of every behavior and associated time stamp. The rationale was that by dividing the gameplay into sections of same puzzles or one to two minutes ranges, the overall agreement can be assessed by comparing the number of matching codes within the section, without granularly looking at every time stamp. Particularly, since when looking at time stamps it was hard to differentiate whether the coders were referring to the exact same instance. Moreover, the inter-rater agreement values fell into the moderate range, ranging from 62% to 78%. Inter-rater agreements dealing with many variables, in this case 25 codes, can be time consuming and require rigorous iterations to ensure high levels of agreements. One observation was that several disagreements were arising from the difference in frequency of a particular behavioral marker, rather than a disagreement on its existence or not. This proved to be challenging in fast paced games, such as simulation genre, where behaviors are happening at a very fast pace, from up to four individuals, therefore creating inconsistencies in the frequencies of behaviors. Therefore, in high paced games, there is more room for errors and missing events. Furthermore, teamwork measurement system should go through several tests to obtain extensive evidence for validity, which is limited in this study.

Selective sampling was followed in this study, since gameplay footage had to meet specific criteria to be coded (e.g., clear commentary, observable in-game characters). Additionally, it can be argued that the sample represents a specific population of individuals who post on YouTube and already have experience in gaming. Therefore, future work will target more representative samples, with teams of a variety of gaming expertise, including novice and non-gamers, to verify whether the cooperative games will still induce the observed behaviors, and how consistency can be affected in these cases. Furthermore, the measurement system has limitation on triangulation, and further work where self-report and automatic measures are implemented can expand on the validity and reliability of the measure.

Third, the process of annotating gameplay footage to track behaviors and associated cooperative features is a time-consuming process, which can limit its wide applicability in future studies. Teamwork studies need dynamic, unobtrusive, real-time measures that can also be done quickly and iteratively. Future work will be needed to automate the coding process. The future of teamwork is more oriented toward automated measures (Roberts et al., 2022). However, observational teamwork measures remain foundational in understanding teamwork behaviors, and even automated measures are still in need of a certain level of human processing to derive insights that cannot be captured with current automation. Therefore, future work that can automate behavioral markers' systems can make the process faster and increase its objectivity and replicability. For example, automated measures have been successfully employed in teamwork communication analysis, where the frequency, the content, and the sequential flow of information can be captured (Marlow et al., 2018; Stanton & Roberts 2020).

Future work aiming to use the codebook of behavioral markers would require additional evidence for the validity of the measurement system, through three validity elements: content, construct and criterion-based (Wiese et al., 2015). Content validity assesses whether the measure represents the construct holistically. Content validity can be partially supported in this study since the codebook was developed through existing teamwork models from foundational team research. However, subject matter experts can further review the measure to detect construct deficiency and contamination (Wiese et al., 2015). In this context, subject matter experts that can provide critical insights include team science experts, or cooperative gaming experts, including experienced designers and gamers. Second, construct validity should be assessed by evaluating whether the measure quantitatively captures the construct of interest (Wiese et al., 2015). This validity would benefit from further triangulation of data, such as collecting self-report measures and automated measures, to provide a holistic perspective on how well the behavioral markers are quantitatively representing the construct of interest. Finally, criterion-related validity evaluates the correlation of the measure with the outcome of interest. This was partially investigated in this study, by comparing the behavioral markers frequencies (measure) with the performance outcomes (outcome of interest). However, the outcome of interest can be team-based outcomes, such as affective states (e.g., satisfaction), and cognitive states (e.g., mental models, team cognition, trust, teamwork perception). These outcomes and their correlations with the measured behaviors, were not measured in this study and should be further explored in future work.

Additionally, the current study used existing commercially available cooperative games, which can limit the researchers' access to controllable parameters when these games are used as testbeds (Harris, 2019). Using existing cooperative games with naturalistic footage allowed the researchers to access authentic player interactions within fully developed games. It has been suggested in previous work, that using low-fidelity gaming prototypes in lab contexts can distort the complex phenomena of social play and affect the emergent nature of social play interactions (Isbister, 2010). However, using games in lab contexts allows researchers to gain a higher level of control over the game parameters, and is more scientifically trackable with less confounding effects, hence allowing for more generalizable research findings (Harris, 2019). Therefore, future work can aim to synthesize these findings through a framework for designing cooperative gamified environments and developing testbeds where researchers can control parameters and run more flexible studies, while attempting to maintain the authentic nature of social gaming interactions. One potential direction is using an event-based approach, where cooperative game features can be used to trigger teaming behaviors, and a teaming measurement system can be applied to track targeted behavioral markers. This will allow for an unobtrusive and systematic measurement of teamwork, in addition to providing a framework with systematic guidelines on how to trigger teamwork behaviors in cooperative games. A demonstration of the framework application through the design of cooperative testbeds will allow for further experimentation and teamwork studies.

Furthermore, the study only studied teamwork within the cooperative gaming environments, while drawing on design insights for testbed development. However, the study does not empirically generalize beyond cooperative environments, to real life teamwork situations where these design insights can be applied. Some parallelism was discussed between the cooperative genres and real-life situations, such as collaborative gaming, aircrew teamwork, emergency medicine and organizational teamwork. However, the examples were provided as insights on how the study can be further expanded to real world teaming contexts. Further work is needed to investigate the external validity of teamwork in cooperative games and to establish a wider generalizability of the results, and further explore the transfer effects between games and similar real-world contexts.

Finally, since this study targeted teamwork in synthetic task environments, future challenges would include the evolving nature of these technologies, where human-autonomy teaming is an emergent field. As teams engage in synthetic tasks, more room is created for complex teamwork compositions, where multiple humans are teamed with multiple agents. This area creates opportunities for future research, and cooperative games can prove to be valid environments for these purposes, especially that the gaming industry is already incorporating AI companions and human-autonomy teaming has been explored in these environments both commercially and academically (Sepich et al., 2021; Bishop et al., 2020).

In conclusion, this work supports the consistency, associations, and validity of cooperative gaming genres in inducing teamwork behaviors, establishing these environments as testbeds for teamwork measurements that can capture dynamic processes in unobtrusive ways, due to their ability to explicitly trigger observable behaviors. The work provides empirical support for using cooperative games for teamwork measurement and suggests a behavioral marker measurement system through a codebook that can be used to measure teamwork in these environments and explore further teamwork studies in the future. Finally, the work supports teamwork measurement and teamwork testbed development through cooperative video games.

REFERENCES

Articles and Video Games

Adams, E., & Dormans, J. (2012). *Game mechanics: advanced game design*. New Riders.

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, *5*, 1-14.

Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & gaming*, *37*(1), 6-23.

Asteroid Base. Lovers in a Dangerous Spacetime. Game. (September 9, 2015). Asteroid Base.

Bakkes, S. C., Spronck, P. H., & van Lankveld, G. (2012). Player behavioural modelling for video games. *Entertainment Computing*, *3*(3), 71-79.

Baron, A., Rayson, P., & Archer, D. (2009). Word frequency and key word statistics in corpus linguistics. *Anglistik*, 20(1), 41-67.

Belanich, J., Sibley, D. E., & Orvis, K. L. (2004). *Instructional characteristics and motivational features of a PC-based game*. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA.

Bergström, K., Björk, S., & Lundgren, S. (2010, October). Exploring aesthetical gameplay design patterns: camaraderie in four games. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 17-24).

Björk, S., & Holopainen, J. (2005). Games and design patterns. *The game design reader: A rules of play anthology*, 410-437.

Bódi, B. (2021). Can Playfulness Be Designed? Understanding Playful Design through Agency in Astroneer (2019). *Eludamos: Journal for Computer Game Culture*, *12*(1), 39-61.

Bowers, C. A., & Jentsch, F. (2001). 7. Use of commercial, off-the-shelf, simulations for team research. In *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 293-317). Emerald Group Publishing Limited.

Bowers, C. A., Braun, C. C., & Morgan Jr, B. B. (1997). Team workload: Its meaning and measurement. In *Team performance assessment and measurement* (pp. 97-120). Psychology Press.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. sage.

Boyle, E., Kennedy, A. M., Doherty, E., O'Keeffe, D., & Traynor, O. (2011). Coping with stress in surgery: the difficulty of measuring non-technical skills. *Irish journal of medical science*, *180*, 215-220.

Braun, M. T., & Kuljanin, G. (2015). Big data and the challenge of construct validity. *Industrial and Organizational Psychology*, *8*(4), 521-527.

Brannick, M. T., Prince, C., & Salas, E. (2005). Can PC-based systems enhance teamwork in the cockpit?. *The International Journal of Aviation Psychology*, *15*(2), 173-187.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: a conceptual analysis and model. *Journal of Applied Psychology*, *91*(6), 1189.

Calleja, G. (2007). Digital game involvement: A conceptual model. *Games and culture*, *2*(3), 236-260.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making.

Carron, A. V., Bray, S. R., & Eys, M. A. (2002). Team cohesion and team success in sport. *Journal of sports sciences*, *20*(2), 119-126.

Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *City*, *1*(2), 1.

Cha, S. H. (2008). Taxonomy of nominal type histogram distance measures. *City*, *1*(2), 1.

Cha, S. H., & Srihari, S. N. (2000, September). Distance between histograms of angular measurements and its application to handwritten character similarity. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 2, pp. 21-24). IEEE.

Chen, M. G. (2009). Communication, coordination, and camaraderie in World of Warcraft. *Games and Culture*, *4*(1), 47-73.

Clever Players. Operation Tango. Game. (August 28, 2020). Clever Plays.

Cohen, J. (1992). Statistical power analysis. *Current directions in psychological science*, *1*(3), 98-101.

Collazos, C. A., Guerrero, L. A., Pino, J. A., Ochoa, S. F., & Stahl, G. (2007). Designing collaborative learning environments using digital games. *J. Univers. Comput. Sci.*, *13*(7), 1022-1032.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, *37*(2), 255-285.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human factors*, *42*(1), 151-173.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, *7*(3), 179.

Cooke, N., Demir, M., & Huang, L. (2020). A framework for human-autonomy team research. In Engineering Psychology and Cognitive Ergonomics. Cognition and Design: *17th International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22 (pp. 134-146).* Springer International Publishing.

Coovert, M. D., Winner, J., Bennett Jr, W., & Howard, D. J. (2017). Serious Games are a Serious Tool for Team Research. *International Journal of Serious Games*, *4*(1).

Costello, B. M. (2018). The rhythm of game interactions: player experience and rhythm in minecraft and don't starve. *Games and Culture*, *13*(8), 807-824.

Crawford, E. R., & Lepine, J. A. (2013). A configural theory of team processes: Accounting for the structure of taskwork and teamwork. *Academy of Management Review*, *38*(1), 32-48.

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field methods*, *23*(2), 136-155.

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3-12.

Dennis, A. R., Robert Jr, L. P., Curtis, A. M., Kowalczyk, S. T., & Hasty, B. K. (2012). Research note—trust is in the eye of the beholder: A vignette study of postevent behavioral controls' effects on individual trust in virtual teams. *Information systems research*, *23*(2), 546-558.

Depping, A. E., & Mandryk, R. L. (2017, October). Cooperation and interdependence: How multiplayer games increase social closeness. In *Proceedings of the Annual Symposium on computer-human interaction in play* (pp. 449-461).

Depping, A. E., Johanson, C., & Mandryk, R. L. (2018, October). Designing for friendship: Modeling properties of play, in-game social capital, and psychological well-being. In *Proceedings of the 2018 Annual Symposium on computer-human interaction in play* (pp. 87-100).

Depping, A. E., Mandryk, R. L., Johanson, C., Bowey, J. T., & Thomson, S. C. (2016, October). Trust me: social games are better than social icebreakers at building trust. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 116-129).

Dormer, R., Cacali, E., & Senna, M. (2017). Having a blast with a computer-mediated information gap task: Keep talking & nobody explodes in the EFL. *Language Teacher*, *41*, 31.

Druskat, V. U., & Kayes, D. C. (1999). The antecedents of team competence: Toward a fine-grained model of self-managing team effectiveness. *Research on managing groups and teams*, *2*(2), 201-231.

Dyer, J. L. (1984). Team research and team training: A state-of-the-art review. *Human factors review*, *26*, 285-323.

Ellis, J. B., Luther, K., Bessiere, K., & Kellogg, W. A. (2008, February). Games for virtual team building. In *Proceedings of the 7th ACM conference on Designing interactive systems* (pp. 295-304).

Emmerich, K., & Masuch, M. (2017, October). The impact of game patterns on player experience and social interaction in co-located multiplayer games. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 411-422).

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human factors*, *37*(1), 65-84.

Farah, Y. A., Dorneich, M. C., & Gilbert, S. B. (2022, September). Evaluating Team Metrics in Cooperative Video Games. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 66, No. 1, pp. 70-74). Sage CA: Los Angeles, CA: SAGE Publications.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, *5*(1), 80-92.

Fernandez, R., Kozlowski, S. W., Shapiro, M. J., & Salas, E. (2008). Toward a definition of teamwork in emergency medicine. *Academic Emergency Medicine*, *15*(11), 1104-1112.

Fullerton, T. (2014). *Game design workshop: a playcentric approach to creating innovative games*. CRC press.

Fine, L. (2016). The Presence of Timers and their Impact on Team Communications During High-Stress Scenarios.

Fletcher, J. D., Tobias, S., & Albay, S. (2006). Using computer games and simulations for instruction: A research review Acquirede Wisdom. Lessons Learned by Distinguished Researchers View project Educational Technology View project. https://www.researchgate.net/publication/228511464

Forzenbyte. 2019. Trine 4: The Nightmare Prince. Game. (October 8, 2019). Modus Games.

Fowlkes, J. E., Lane, N. E., Salas, E., Franz, T., & Oser, R. (1994). Improving the measurement of team performance: The TARGETs methodology. *Military Psychology*, *6*(1), 47-61.

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The international journal of aviation psychology*, *8*(3), 209-221.

Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. *Handbook of industrial and organizational psychology*, *4*(2), 271-340.

Goebel, R., Wahlster, W., & Siekmann, J. (n.d.). Lecture Notes in Artificial Intelligence Subseries of Lecture Notes in Computer Science Series Editors Founding Editor. http://www.springer.com/series/1244

Goodwin, G. F., Blacksmith, N., & Coats, M. R. (2018). The science of teams in the military: Contributions from over 60 years of research. *American Psychologist*, *73*(4), 322.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2017). Measuring team situation awareness in decentralized command and control environments. In *Situational Awareness* (pp. 183-196). Routledge.

Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Detecting perturbations to team communication. *Ergonomics*, *55*(8), 825-839.

Granåsen, M. (2019). *Exploring C2 Capability and Effectiveness in Challenging Situations: Interorganizational Crisis Management, Military Operations and Cyber Defence* (Vol. 1836). Linköping University Electronic Press.

Grand, J. A., Pearce, M., Rench, T. A., Chao, G. T., Fernandez, R., & Kozlowski, S. W. (2013). Going DEEP: guidelines for building simulation-based team assessments. *BMJ quality & safety*, *22*(5), 436-448.

Grandi, J. (2021). Fast communication training through asymmetrical multiplayer videogames.

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *psychological science in the public interest*, *19*(2), 59-92.

Gries, S. T. (2010). Useful statistics for corpus linguistics. A mosaic of corpus linguistics: Selected approaches, 66, 269-291.

Guenaga, M., Eguíluz, A., Rayón, A., Núñez, A., & Quevedo, E. (2014, November). A serious game to develop and assess teamwork competency. In *2014 International Symposium on Computers in Education (SIIE)* (pp. 183-188). IEEE.

Halbrook, Y. J., O'Donnell, A. T., & Msetfi, R. M. (2019). When and how video games can be good: A review of the positive effects of video games on well-being. *Perspectives on Psychological Science*, *14*(6), 1096-1104.

Hämäläinen, R. H., Niilo-Rämä, M., Lainema, T., & Oksanen, K. (2018). How to raise different game collaboration activities: The association between game mechanics, players' roles and collaboration processes. *Simulation & Gaming*, *49*(1), 50-71.

Hämäläinen, R., & Vähäsantanen, K. (2011). Theoretical and pedagogical perspectives on orchestrating creativity and collaborative learning. *Educational Research Review*, *6*(3), 169-184.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

Handler, E. (2017, October). Games: Facilitating Communications Training (and Fun). In *Proceedings of the 2017 ACM SIGUCCS Annual Conference* (pp. 127-128).

Harris, J. J. (2019). Leveraging asymmetry and interdependence to enhance social connectedness in cooperative digital games.

Harris, J., Hancock, M., & Scott, S. D. (2016, October). Leveraging asymmetries in multiplayer games: Investigating design elements of interdependent play. In *Proceedings of the 2016 Annual Symposium on computer-human interaction in play* (pp. 350-361).

Hazelight Studios. 2021. It Takes Two. Game. (March 25, 2021). Electronic Arts.

Healey, A. N., Undre, S., & Vincent, C. A. (2004). Developing observational measures of performance in surgical teams. *BMJ Quality & Safety*, *13*(suppl 1), i33-i40.

Hooshyar, D., Yousefi, M., & Lim, H. (2018). Data-driven approaches to game player modeling: a systematic literature review. *ACM Computing Surveys (CSUR)*, *50*(6), 1-19.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004, July). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, No. 1, p. 1722).

Hussain, T. S., Weil, S. A., Brunyé, T., Sidman, J., Ferguson, W., & Alexander, A. L. (2008). Eliciting and evaluating teamwork within a multi-player game-based training environment. *Computer games and team and individual learning*, 77-104.

Isbister, K. (2010). Enabling social play: A framework for design and evaluation. *Evaluating user experience in games: Concepts and methods*, 11-22.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annu. Rev. Psychol.*, *56*, 517-543.

Indoor Astronaut. Unrailed. Game. (September 9, 2019). Daedalic Entertainment.

Inertia Game Studios. Catastronauts. Game. (September 28, 2018). Intertia Game Studios.

Jang, Y., & Ryu, S. (2011). Exploring game experiences and game leadership in massively multiplayer online role-playing games. *British Journal of Educational Technology*, *42*(4), 616-623.

Janz, B. D., Colquitt, J. A., & Noe, R. A. (1997). Knowledge worker team effectiveness: The role of autonomy, interdependence, team development, and contextual support variables. *Personnel psychology*, *50*(4), 877-904.

Jentsch, F., & Bowers, C. A. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *The International Journal of Aviation Psychology*, *8*(3), 243-260.

Kendall, D. L., & Salas, E. (2004). Measuring team performance: Review of current methods and consideration of future needs. *The science and simulation of human performance*, *5*, 307-326.

Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J. C., & Martin, M. J. (2002, September). Some promising results of communication-based automatic measures of team cognition. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, No. 3, pp. 298-302). Sage CA: Los Angeles, CA: SAGE Publications..

Klei Entertainment. Don't Starve Together. Game. (April 21, 2016). Klei Entertainment.

Klong Games. 2015. Shift Happens. Game. (September 9, 2015). Klong Games.

Kozlowski, S. W., Gully, S. M., Nason, E. R., & Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. *Pulakos (Eds.), The changing nature of work performance: Implications for staffing, personnel actions, and development*, *240*, 292.

Kutlu, B., Bozanta, A., & Nowlan, N. (2013, July). Multi-user virtual environments and serious games for team building in organizations. In *Proceedings of the sixth international conference on e-learning in the workplace (ICELW 2013). New York, USA*.

Levy, L. M., Solomon, R., Johnson, J., Wilson, J., Lambeth, A. J., Gandy, M., ... & Liu, R. (2016). Grouches, Extraverts, and Jellyfish: Assessment Validity and Game Mechanics in a Gamified Assessment. In *DiGRA/FDG*.

Lisk, T. C., Kaplancali, U. T., & Riggio, R. E. (2012). Leadership in multiplayer online gaming environments. *Simulation & Gaming*, *43*(1), 133-149.

Long, C. A. (2008). Recent advances on applied mathematics: proceedings of the American Conference on Applied Mathematics (MATH '08) : Cambridge, Massachusetts, USA, March 24-26, 2008. WSEAS Press.

Mariais, C., Michau, F., & Pernin, J. P. (2012). A description grid to support the design of learning role-play games. *Simulation & Gaming*, *43*(1), 23-33.

Marks, M. A., & Panzer, F. J. (2004). The influence of team monitoring on team processes and performance. *Human Performance*, *17*(1), 25-41.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of management review*, *26*(3), 356-376.

Marlow, S. L., Salas, E., Landon, L. B., & Presnell, B. (2016). Eliciting teamwork with game attributes: A systematic review and research agenda. *Computers in Human Behavior*, *55*, 413-423.

Marlow, S., Bisbey, T., Lacerenza, C., & Salas, E. (2018). Performance measures for health care teams: A review. *Small Group Research*, *49*(3), 306-356.

Mayer, I. (2018). Assessment of teams in a digital game environment. *Simulation & Gaming*, *49*(6), 602-619.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing team performance: Emerging principles from complex environments. *Team effectiveness and decision making in organizations*, *16*, 9-45.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., Frezzo, D. C., & West, P. (2012). Three things game designers need to know about assessment. *Assessment in game-based learning: Foundations, innovations, and perspectives*, 59-81.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.

Morgan Jr, B. B., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *The Journal of General Psychology*, *120*(3), 277-291.

Morschheuser, B., Maedche, A., & Walter, D. (2017, February). Designing cooperative gamification: Conceptualization and prototypical implementation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 2410-2421).

Morschheuser, B., Riar, M., Hamari, J., & Maedche, A. (2017). How games induce cooperation? A study on the relationship between game features and we-intentions in an augmented reality game. *Computers in human behavior*, *77*, 169-183.

Musick, G., Zhang, R., McNeese, N. J., Freeman, G., & Hridi, A. P. (2021). Leveling up teamwork in esports: Understanding team cognition in a dynamic virtual environment. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-30.

Nasir, M., Lyons, K., Leung, R., Bailie, A., & Whitmarsh, F. (2015, November). The effect of a collaborative game on group work. In *Proceedings of the 25th annual international conference on computer science and software engineering* (pp. 130-139).

NExT Studios. 2020. BiPed. Game. (March 26, 2020). NExT Studios, META publishing, PostMeta Games Limited.

Oakes, M. (2019). *Statistics for corpus linguistics*. Edinburgh University Press.

Obsidian Entertainment. Grounded. Game. (July 28, 2020). Xbox Game Studios.

Other Tales Interactive. Tick Tock: A Tale for Two. Game. (February 28, 2019). Other Tales Interactive.

Ouverson, K. M., & Gilbert, S. B. (2021). A composite framework of co-located asymmetric virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-20.

Oulasvirta, A. (2009). Social inference through technology. *Awareness Systems: Advances in Theory, Methodology and Design*, 125-147.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw-hill education (UK).

Peña, J., & Hancock, J. T. (2006). An analysis of socioemotional and task communication in online multiplayer video games. *Communication research*, *33*(1), 92-109.

Peppler, K., Danish, J. A., & Phelps, D. (2013). Collaborative gaming: Teaching children about complex systems and collective behavior. *Simulation & Gaming*, *44*(5), 683-705.

Pinelle, D., Gutwin, C., & Greenberg, S. (2003). Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *10*(4), 281-311.

Pobiedina, N., Neidhardt, J., Moreno, M. D. C. C., Grad-Gyenge, L., & Werthner, H. (2013, July). On successful team formation: Statistical analysis of a multiplayer online game. In *2013 IEEE 15th Conference on Business Informatics* (pp. 55-62). IEEE.

Rico, R., Sánchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *Academy of management review*, *33*(1), 163-184.

Rasker, P. C., Post, W. M., & Schraagen, J. M. C. (2000). Effects of two types of intra-team feedback on developing a shared mental model in Command & Control teams. *Ergonomics*, *43*(8), 1167-1189.

Reuter, C., Wendel, V., Göbel, S., & Steinmetz, R. (2014, August). Game Design Patterns for Collaborative Player Interactions. In *DiGRA*.

Robert Jr, L. P. (2016, February). Monitoring and trust in virtual teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 245-259).

Roberts, A. P., Webster, L. V., Salmon, P. M., Flin, R., Salas, E., Cooke, N. J., ... & Stanton, N. A. (2022). State of science: models and methods for understanding and enhancing teams and teamwork in complex sociotechnical systems. *Ergonomics*, *65*(2), 161-187.

Roberts, K., Dowell, A., & Nie, J. B. (2019). Attempting rigour and replicability in thematic analysis of qualitative research data; a case study of codebook development. *BMC medical research methodology*, *19*, 1-8.

Rocha, J. B., Mascarenhas, S., & Prada, R. (2008). Game mechanics for cooperative games. *ZON Digital Games 2008*, 72-80.

Rosen, M. A., Bedwell, W. L., Wildman, J. L., Fritzsche, B. A., Salas, E., & Burke, C. S. (2011). Managing adaptive performance in teams: Guiding principles and behavioral markers for measurement. *Human resource management review*, *21*(2), 107-122.

Rosen, M. A., Salas, E., Wu, T. S., Silvestri, S., Lazzara, E. H., Lyons, R., ... & King, H. B. (2008). Promoting teamwork: An event-based approach to simulation-based teamwork training for emergency medicine residents. *Academic Emergency Medicine*, *15*(11), 1190-1198.

Rosen, M. A., Schiebel, N., Salas, E., Wu, T. S., Silvestri, S., & King, H. B. (2012). How can team performance be measured, assessed, and diagnosed. *Improving patient safety through teamwork and team training*, 59-79.

Rosenfeld, L. B., & Richman, J. M. (1997). Developing effective social support: Team building and the social support process. *Journal of Applied Sport Psychology*, *9*(1), 133-153.

Rousseau, V., Aubé, C., & Savoie, A. (2006). Teamwork behaviors: A review and an integration of frameworks. *Small group research*, *37*(5), 540-570.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, *50*(3), 540-547.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training.

Salas, E., Reyes, D. L., & Woods, A. L. (2017). The assessment of team performance: Observations and needs. *Innovative assessment of collaboration*, 21-36.

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, space, and environmental medicine*, *78*(5), B77-B85.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human resource management*, *54*(4), 599-622.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “big five” in teamwork?. *Small group research*, *36*(5), 555-599.

Seif El-Nasr, M., Aghabeigi, B., Milam, D., Erfani, M., Lameman, B., Maygoli, H., & Mah, S. (2010, April). Understanding and evaluating cooperative games. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 253-262).

Sicart, M. (2015, June). Loops and Metagames: Understanding Game Design Structures. In *FDG*.

Sottilare, R. A., Shawn Burke, C., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, *28*, 225-264.

Steel Crate Games. Keep Talking and Nobody Explodes. Game. (July 16, 2015). Steel Crate Games.

Stonewheat & Sons. KeyWe. Game. (June 16, 2020). Fireshrine Games.

System Era Softworks. Astroneer. Game. (December 16, 2019). Gearbox Publishing.

Tang, A., Tory, M., Po, B., Neumann, P., & Carpendale, S. (2006, April). Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1181-1190).

Team 17 & Ghost Town Games. Overcooked 2. (August 7, 2018). Team 17.

Team 17 & Mouldy Touf Studios. The Survivalists. Game. (October 9, 2020). Team 17, Fireshrine Games.

Teruel, M. A., Navarro, E., González, P., López-Jaquero, V., & Montero, F. (2016). Applying thematic analysis to define an awareness interpretation for collaborative computer games. *Information and Software Technology*, *74*, 17-44.

Minkkinen, T. (2016). Basics of Platform Games.

Total Mayhem Games & TMG Studios B.V. We Were Here Forever. Game. (2022). Total Mayhem Games.

Toups, Z. O., Hammer, J., Hamilton, W. A., Jarrah, A., Graves, W., & Garretson, O. (2014, October). A framework for cooperative communication game mechanics from grounded theory. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (pp. 257-266).

Vaddi, D., Wehbe, R. R., Toups, Z. O., Stahlke, S. N., Koroluk, R., & Nacke, L. E. (2015, October). Validating test chambers to study cooperative communication mechanics in Portal 2. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 721-726).

Valve Corporation. 2011. Portal 2. Game. (April 19, 2011). Valve Corporation.

Vargas-Iglesias, J. J. (2020). Making sense of genre: The logic of video game genre organization. *Games and Culture*, *15*(2), 158-178.

Wang, Q. (2009). Design and evaluation of a collaborative learning environment. *Computers & Education*, *53*(4), 1138-1146.

Weaver, S. J., Rosen, M. A., Salas, E., Baum, K. D., & King, H. B. (2010). Integrating the science of team training: guidelines for continuing education. *Journal of Continuing Education in the Health Professions*, *30*(4), 208-220.

Webber, S. S. (2002). Leadership and trust facilitating cross-functional team success. *Journal of management development*.

Wendel, V., & Konert, J. (2016). Multiplayer serious games. *Serious Games: Foundations, Concepts and Practice*, 211-241.

Whittaker, G., Abboudi, H., Khan, M. S., Dasgupta, P., & Ahmed, K. (2015). Teamwork assessment tools in modern surgical practice: a systematic review. *Surgery research and practice*, *2015*.

Wickens, T. D. (Ed.). (2014). *Multiway contingency tables analysis for the social sciences*. Psychology Press.

Wiese, C. W., Shuffler, M. L., & Salas, E. (2015). Teamwork and team performance measurement.

Williams, J. P., & Kirschner, D. (2012). Coordinated action in the massively multiplayer online game World of Warcraft. *Symbolic Interaction*, *35*(3), 340-367.

Wilson, D. W., Jenkins, J., Twyman, N., Jensen, M., Valacich, J., Dunbar, N., ... & Nunamaker, J. F. (2016, January). Serious games: an evaluation framework and case study. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 638-647). IEEE.

Wittenbaum, G. M., Hollingshead, A. B., Paulus, P. B., Hirokawa, R. Y., Ancona, D. G., Peterson, R. S., ... & Yoon, K. (2004). The functional perspective as a lens for understanding groups. *Small group research*, *35*(1), 17-43.

Wuertz, J., Alharthi, S. A., Hamilton, W. A., Bateman, S., Gutwin, C., Tang, A., ... & Hammer, J. (2018, April). A design framework for awareness cues in distributed multiplayer games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Yeatts, D. E., & Hyten, C. (1998). *High-performing self-managed work teams: A comparison of theory to practice*. Sage.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical education*, *40*(11), 1098-1104.

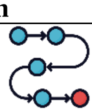
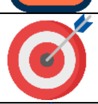
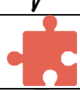
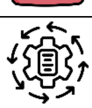
Zaccaro, S. J., Marks, M., O'Connor-Boes, J., & Costanza, D. (1995). The nature and assessment of leader mental models. *Management Research Institute, Bethesda, MD (Tech. Rep. MRI 95-3).*

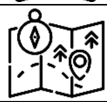
Zagal, J. P., Rick, J., & Hsi, I. (2006). Collaborative games: Lessons learned from board games. *Simulation & gaming*, *37*(1), 24-40.

Zea, N. P., Sánchez, J. L. G., Gutiérrez, F. L., Cabrera, M. J., & Paderewski, P. (2009). Design of educational multiplayer videogames: A vision from collaborative learning. *Advances in Engineering Software*, *40*(12), 1251-1260.

Zhang, X., Venkatesh, V., & Brown, S. A. (2011). Designing collaborative systems to enhance team performance. *Journal of the Association for Information Systems*, *12*(8), 556-584.

Visual Summary Icons

Icon	Attribution
	Icon made by Becros from www.flaticon.com
	Icon made by bqlqn from www.flaticon.com
	Icon made by Freepik from www.flaticon.com
	Icon made by Smashicon from www.flaticon.com
	Icon made by Freepik from www.flaticon.com
	Icon made by DinosoftLabs from www.flaticon.com
	Icon made by juicy_fish from www.flaticon.com
	Icon made by Freepik from www.flaticon.com
	Icon made by Designing Hub from www.flaticon.com
	Icon made by Freepik from www.flaticon.com

	Icon made by Nhor Phai from www.flaticon.com
	Icon made by itim2101 from www.flaticon.com
	Icon made by Ayub Irawan from www.flaticon.com
	Icon made by Freepik from www.flaticon.com
	Icon made by Freepik from www.flaticon.com

APPENDIX A. CODEBOOK OF BEHAVIORAL MARKERS AND COOPERATIVE FEATURES

Behavioral Marker	Definition	Description in Games	Qualifications	Examples	Code
		Situation Assessment			
Cue recognition	A team member (or more) scanning the environment for cues that can influence the mission. (Marks et al., 2001; Rosen et al., 2011)	A team member's avatar is visibly seen scanning the environment (looking around, scrolling the screen) as an initial situation assessment to detect cues. Verbal expressions can be associated with the behaviors that give signs that the member is scanning or looking for cues, such as relaying information that they see cues ("I see a button")	-To be behaviorally seen or verbally heard -It must be in the transition process (hence, it is coded when the team engage in this scanning when they first encounter the task or the mission and still figuring out how to proceed)	-"There's a door there" -"I see a button next to the staircase"	SA-CR
Cue meaning	After scanning the environment, team members connect how the cues can affect the mission by interpreting this information (Marks et al., 2001; Rosen et al., 2011)	After scanning, team members start figuring out how the recognized cues can affect their mission	-Cue meaning should be verbally expressed -Cue meaning is not Problem Solving but rather an assessment of the cue's functions -It is a transition process and therefore is coded when members first encounter the obstacle or when they're understanding how to proceed	-"I think the button can open the door" -"Maybe the lever can lower the staircase"	SA-CM
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
		Analysis and Planning			
Mission Analysis	Team members engaging in discussions on their purpose in the specific task/mission, how the resources and information can be applied to meet goals and what approaches can contribute to their advancement.	Team members having collaborative conversations involving analyzing the gameplay mission, understanding the goal of the level, discussing potential approaches.	-Mission analysis is a team level process where more than one team member should be involved in the discussion -It should be verbally heard -It is a transition process. And therefore, it is coded when team members are understanding their mission. -Back and forth	"Maybe we should attach the nail to the wall; Oh I can swing over it?" "I think we go crazy in darkness; what if we build a base camp before that? Oh we can build a fire?"	MA

			collection of mission relevant information		
Deliberate Planning	The process of brainstorming plans and strategies and explicitly formulating a course of action to take (Marks et al., 2001; Rosen et al., 2011)	Deliberate planning can happen fast in games however it should be captured when the team decided on a course of action on how to proceed.	-Verbally expressed -A formulation of the next steps to do (Plan A) -It is a projection of future steps that has been verbally expressed	-"Okay let's swing the hammer through the nail to get to the other side" -"We should first build a base camp and tomorrow we'll go look for resources"	DP
Team Leadership Behaviors	The processes of team leadership: directing the team, taking charge in developing plans, formulating objectives, taking lead in guiding the team in a mission (Salas et al., 2005)	In cooperative games, leadership can be observed when a team-mate or more take lead depending on the task/mission. It is observed when one player is clearly taking charge or assuming leadership roles, such as commanding other players to do a certain action.	-It must be verbally heard -It must be addressing the whole team		TL
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Explicit Coordination					
Synchronizing or Sequencing	Team members pacing their activities through verbally sequencing tasks or synchronizing movements.	In Games, team members sequence tasks through assigning task roles in a certain order to execute interdependent game activities. Team members synchronize when their movements through timing. -Player A providing information to sequence Player B's actions -Player A and B provide information to sequence each other's actions	-It has to be verbally heard -Sequencing or synchronizing here is an action process and therefore it's not coded when it's a part of the plan formulation.	-"1,2,3, go!" -"I'll get the wood first and you can add the grass"	EC-S
Reporting	Team members relaying messages about their standing, requirements and goals	When team members report to the team about their status (e.g., location), needs (e.g., food) or objectives (e.g., what they're doing next), or observations.	-It must be verbally heard	-"I'm next to the river now" -"I need some food" -"I'm chopping trees"	EC-R
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Implicit Coordination (Rico et al., 2008)					
Sequencing or synchronizing	When Team-mates sequence or synchronize their actions (Marks et al., 2001) (same functions as	In a sequential task, when players do a game task in sequence or synchrony without communicating (Wuertz et al., 2018), or when synchronizing actions (e.g.,	-It must be behaviorally observable (therefore an action or a movement by the player avatar or	-Team mates jumping on two buttons at the same time without communicating -Team mates	IC-S

	explicit coordination) however without explicitly communicating	jumping on buttons), they do it without communicating explicitly.	character)	ordering their ingredients (e.g., Overcooked) without communicating	
Anticipation	When team-mates anticipate the need s of the task or the team without prior communication or planning	When a team-mate does an action that contributes to the team's progress (e.g., building a fire in the base camp; preparing food in advance) without prior communication or explicit requests.	-It must be behaviorally observable or at least verbally communicated after the action is done (e.g., I built a fire). In which case, it will be coded IC-A (the action of building) and EC-R (the action of reporting) -It must be an action that serves the team or a team-mate or more rather than an individual action (e.g., eating)	-As mentioned in the second column, building a fire, or working in the base camp (e.g., Don't Starve) without being told to do it -Repairing ship damage that affects the collective health status	IC-A
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Mutual Performance Monitoring					
Explicitly verbal	Explicit performance monitoring is when a team mate verbally expresses their monitoring toward other team-mates (Salas et al., 2005; Marks & Panzer, 2004; McIntyre & Salas, 1995; Rosen et al., 2001)	For the sake of coding, explicit monitoring is whenever a team-mate asks a question or does a verbal indicator that they're tracking other team-mates' performance	-It must be verbally expressed -It must be team members checking on other team members' performance status	- "How are you doing on food?" - "Where are you now?" - "Your sanity level is low"	MPM-E
Explicit through verbal	MPM can be a cognitive process and therefore it might be happening cognitively in a way that cannot be captured. However, it is whenever teammates are tracking other's performance or status without verbally asking.	For the sake of coding, implicit MPM is whenever team-mates give verbal or behavioral signs that they were doing MPM implicitly (e.g., stopping to observe the team-mates or check on them; or offering help even if it wasn't asked)	-It has to be implicitly performed however captured through verbal or behavioral signs -It has to be teammates monitoring other teammates performance	Since MPM cannot be verbally heard it will be coded in two cases: -Behaviorally (e.g., teammates looking at other's performance through their avatar) -Or when it's associated with some type of backup (feedback, behavior, resources) that	MPM-E

				was not requested (e.g., "Here's some food"; "You could try jumping higher".	

Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Backup Behavior (Salas et al., 2005; Rousseau et al., 2006; Marks et al., 2001; Kozlowski et al., 1999; Rosen et al. 2011)					
Proactive Backup Behavior	When team members proactively assist others (Rosen et al., 2011); the helper initiate the backup action	When team-mates provide others with feedback, resources, or task assistance without being asked to, after recognizing a need.	-It has to be verbally or behaviorally detected -It has to be initiated by the help provided (proactive action) -Proactive backup might be preceded by mutual monitoring (watch out for it)	-Providing food for a starving team-mate -Providing feedback on how surpass an obstacle in the game	PBB-
Reactive Backup Behavior	Reactive backup is when team-mates who need assistance initiate the request by indicating that they need help	When team-mates report a need and a team-mate respond through helping; or when team-mates directly request for task assistance or performance guidance.	-It has to be verbally or behaviorally detected -It has to be initiated by the person receiving help (through verbal signs or direct requests/questions)	- "I'm starving"; "Here's some food" - "I'm being attacked now", "I'm coming to help"	RBB-
Backup Behavior types (Associated with PBB or RBB)					
Sub-type: Feedback	A type of backup (PBB or RBB) where teammates provide feedback to assist another team-mate in performance or provide guidance	Team-mates telling others how to overcome an obstacle if they're struggling or failing to do so. Or answering their questions.	-It has to be verbally detected -It has to be feedback generated from a need	(A team-mate failing to jump across the platform) "You should jump vertically and then move forward" "How did you do that?", "I made an axe"	RBB-F PBB-F
Sub-type: Behavioral assistance	When a team-mate come and help another by assisting them through the task or doing the task themselves (if a need is recognized).	Team-mates behaviorally showing others how to do something or assisting them with their tasks by redistributing the load	-This is not a feedback backup and therefore it should be a behavioral assistance -It is a backup IF it is assisting in a need (a team-mate overloaded, lacking behind, attacked, failing to do	-a team-mate joining an overloaded team-mate on their station in overcooked (the cooking game) -team-mate clearing the obstacle again so their other team-mates can see how to do it	RBB-B PBB-B

			something…)		
Sub-type: Resources	Teammates sharing their resources with another teammate when a need is recognized	In a game where "sharing abilities" are available, this form of backup is when a teammate gives from their own resources to another player.	-This is a resource sharing backup (giving food, tools, equipment to another player) -This would be observable in a game where resources can be shared. -This is directed toward another player to support them through resources when needed	-A teammate giving food to another teammate -A teammate giving equipment to another teammate (As a response to a need)	PBB-R RBB-R
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Systems Monitoring					
Environmental Monitoring	Tracking environmental conditions and resources that are relevant to mission accomplishment (Marks et al., 2001)	Team members tracking the game environment conditions (Teruel et al., 2016) (for resources, risks, conditions, new puzzles, or obstacles)	-It has to be behaviorally or verbally detected -It has to be directed toward external environmental conditions -Environmental conditions are not cues (be careful not to mix it with situation assessment)	-Team members looking around the environment for any new resources (e.g., trees, grass) -Diffuser flipping the bomb to see the number of batteries (keep talking) -Team-members keeping track of time before it runs out	SM-E
Internal Monitoring	Tracking the internal team environment, including team resources or generated/acquired information (Marks et al., 2001)	Team members asking or checking the team's inventory/equipment/tools; Team members checking what other team-mates currently have. Team-mates asking or revisiting information that the team has discovered.	-It has to be behaviorally or verbally detected -It has to be directed toward the internal team resources (and not performance)	-Team members asking how much food they still have in the inventory -Team members checking what tools a team member currently has	SM-I
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Adaptive Behaviors					
Reactive Strategy Adjustment (Ilgen et al., 2005)	Team members adjusting their formulated plan, or their current course of action based on their performance advancement or due to changes in	When team members are already engaged in a task based on deliberate planning or initial assessment/analysis of the mission however they adjust their strategies as they go	-This is an action phase and therefore must be performed by team members while they are executing or performing a task	-Team members already trying a certain strategy to solve a puzzle however discover it's not working and therefore develop a new way to solve it	RAS

	the environment		-It must be an adaptive adjustment -It must be verbally heard -It must be a team level adjustment		
Team Learning: Review performance (Ilgen et al., 2005)	In this part, team members are learning from their performance, interpreting how their actions are affecting that and adjusting accordingly.	Team members perceiving cues/information from their performance and notice errors and consequences	-It can be happening as the team perform and learn as they go -It can be happening after performance as a preparation for the next phase	-Team members noticing that what they're doing is not working and therefore revise their knowledge/performanc e -Team members reflecting on their past performance and reviewing what worked and didn't work for the next task	TL-R
Behavioral Marker	**Definition**	**Description in Games**	**Qualifications**	**Examples**	**Code**
Cohesion and Social					
Team Members encouragemen t (Sottilare et al., 2018)	When team members encouraging each through acknowledging each other's strength or complementing weaknesses	In the context of games, it happens verbally through team-mates encouraging each other's or the team.	-Verbally expressed -Should be directed toward team members and not their taskwork or teamwork	-"We're a great team" -"Great job" -"Yes you did it!"	TC-E
Complimenting team skills (Sottilare et al., 2018)	Team members expressing compliments toward the teams' taskwork or teamwork	When team members comment on their task performance or teamwork skills	-Verbally expressed -Oriented toward the team processes, tasks, and skills	-"great teamwork" -"We're so good at this"	TC-C
Social interactions	Team members having casual conversations or joking		-Verbally heard -Not task oriented but rather social/casual (e.g., talking about the game story, talking about other topics) -This excludes personal stories that are irrelevant to the game that we do not code		SI

Feature	Description	Code
Gameplay Progression Features		
Shared Obstacle (Seif El Nasr et al., 2010; Manninen & Korva, 2005)	A challenge that requires only one action to be cleared. A shared obstacle must be cleared by at least more than one team member. A shared obstacle is either an obstacle that requires an action where all team-mates have a role (for example pressing two buttons to open a door) or that is needed to be cleared in order for the team to advance (for example one player lowering the staircase so everyone else can pass)	SO
Shared Puzzle (Seif El Nasr et al., 2010)	A Challenge that requires problem solving and is associated with a set of actions in order to be cleared. A shared puzzle requires a series of actions directed toward the same goal. In some contexts, a shared puzzle can be a series of shared obstacles that have the same end goal. In this case it is coded as shared puzzle (for example the following series of actions: elevating the pipe, passing through the pipe, pressing the button, jumping above the fan)	SP
Complementary Obstacle (Zagal et al., 2008; Seif El Nasr et al., 2010)	A challenge that requires only one action to be cleared. Additionally, players are equipped with complementary abilities. Complementary abilities are equipped from the game and not the same as complementary roles in a puzzle (for example in a shared puzzle a player elevates the pipe while another presses a button, this is not complementary abilities; while in another puzzle, where a player has a hammer and another has a nail, this is complementary abilities)	CO
Complementary Puzzle (Zagal et al., 2008; Seif El Nasr et al., 2010)	A challenge that requires problem solving and is associated with a set of actions to be cleared. Additionally, players are equipped with complementary abilities.	CP
Asymmetric Obstacle (Harris, 2019)	A challenge that has same description as SO however it involves an asymmetry in challenge, interface, information, or role (Harris, 2019)	AO
Asymmetric Puzzle (Harris, 2019)	A challenge that has same description as SP however involves an asymmetry in challenge, interface, information, or role (Harris, 2019)	AP
Boss Fights	A shared Boss located at the end of levels/chapters. It differs than common Enemies in size, level of challenge and frequency.	BF
Boss Fights Challenges	Puzzles or Obstacles in the Boss fights, where players are required to solve the challenge to defeat the boss fight. Similar in description to SO/SP however implemented in the Boss fight with extra pressures and risks	BFC
Common Enemies	Enemies that attack more than one player.	CE
Individual Obstacle	An obstacle or an enemy faced by one player without the requirement of assistance from other players. These obstacles are cleared on an individual basis. There is no dependence on other players to assist or clear the obstacle for other players to advance. These obstacles must be cleared every time a player encounters them.	IO
Competitive Challenge	Side-quest challenges where players compete against each other's.	CC
Task Allocation Continuous Puzzle	A continuous puzzle in the gameplay level, where players are continuously working on the task that has a shared goal. The puzzle involves task allocation and therefore players can switch	TACP

	tasks and work on interdependent tasks. The TACP constitutes the level and is not a part of a progressive gameplay like in platformer.	
Exploratory Fun Mechanics	Fun mechanics in games for exploratory purposes	EFM
Interactive Shared Object (Seif El Nasr et al., 2010)	An object that can be altered and moved by players (e.g., boxes)	ISO
Player Ability	**Description**	**Code**
	Player Abilities	
Sharing Abilities	Players equipped with the abilities to share their resources with others.	SA
Environment Modifying Ability	Players equipped with abilities to modify their environments (e.g., open portals), for personal reasons (to advance in the game) or for collective reasons (assisting in solving puzzles or obstacles; In this case EMA is paired with SP.	EMA
Limited Life Resources	Limited resources associated with player's life (such as life meter, food meter). This feature is only coded when it is actively imposing risks on the players and motivating their actions (for example out of health)	LLR
Cooperative Communication Mechanics (CCM) (Vaddi et al., 2015)	Communication Mechanics embedded in the game that allow players to communicate in ways other than verbal communication (e.g., ping tools, gestures)	CCM
Crafting Abilities (Sicart et al., 2015)	Abilities where players can build and craft in the environment (Survival Games)	CA
Environmental Component	**Description**	**Code**
	Environmental Components	
Shared Utilities (Seif El Nasr et al., 2011)	Resources that are used by all players and can be already built in in the game or because of shared collectables.	SU
Environment Resources	Resources that can be collected by players and go to their individual inventories. Resources can be used to benefit only the player or the team.	ER
Team Awareness Cues (Wuertz et al., 2018)	Features observed in multiplayer games that allow players to track each other's status (for example Mini-Map, visible health bars)	TAC
MOB Resources	Resources that can interact with the player and impose risks (for example attack the player)	MR
Common Risks	Risks available in the game environment that impose danger on more than one player. Risks are not enemies. (Example: darkness). This feature is coded only when it is actively imposing risks on the players and motivating their actions	CR
Game Pressures	Game pressures such as time limit that impose extra pressures on players to complete their tasks. Coded when it is actively imposing pressures on the players and motivating their actions.	GP
Feature	**Description**	**Code**
	Story, aesthetics and game world	
Story (Schell, 2008)	Plot formation, character dialogues, story themes (Schell, 2008).	S
Aesthetics	Game environment, graphics, character movements and characteristics (Schell, 2008).	A
Shared Environment	The shared environment where players are interacting.	SE
Asymmetric Environment	The asymmetric environment where players are interacting.	AE
Dynamically Changing Environment	An environment that is mechanically coded to dynamically change (e.g., kitchen changing layout in OC2).	DCE
Dynamic	**Description**	**Code**
	Emergent play cooperative dynamics	
Community Survival	A dynamic (observed in survival games) that equips players with higher chances of survival if they build and maintain a community.	(CS) (coded in parenthesis,

		paired with other features)
Team Spirit	A dynamic (observed in survival games) that equips players with higher chances of survival if they ensure the survival of other team-mates	(TS) (coded in parenthesis paired with other features)

APPENDIX B. LIST OF SCREENED COOPERATIVE GAMES

Game #	Game Name	Genre	Shared Goal	Communication	Interdependence	Coupling	Decision
1.	It Takes Two	Action Puzzle Platformer	Yes	Yes	High	Closely	PASS 1
2.	Don't Starve Together	Survival	Yes	Yes	Intermediate	Loosely	PASS 2
3.	Keep Talking and Nobody Explodes	Asymmetric	Yes	Yes	High	Closely	PASS 3
4.	Lovers in A Dangerous Spacetime	Multiplayer Space Shooter	Yes	Yes	Intermediate	Closely	PASS 4
5.	Trine 4	Action Puzzle Platformer	Yes	Yes	High	Closely	PASS 5
6.	Moving Out	Moving Simulation	Yes	Yes	Low/Intermediate	Closely	BACKUP 1
7.	Overcooked 2	Cooking Simulation	Yes	Yes	Intermediate	Closely	PASS 6
8.	SnipperClips	Puzzle	Yes	Yes	High	Closely	PASS 7
9.	Portal 2	Action Puzzle Platformer	Yes	Yes	High	Closely	PASS 8
10.	KeyWe	Puzzle Simulation	Yes	Yes	Intermediate	Closely	PASS 9
11.	Phogs	Puzzle Platformer	Yes	Yes	Intermediate	Closely	PASS 10
12.	For the King	RPG tabletop	Yes	Yes	Low	Closely	FAIL 1
13.	The Survivalists	Survival	Yes	Yes	Intermediate	Loosely	PASS 10
14.	We Were Here Forever	Asymmetric	Yes	Yes	High	Closely	PASS 11
15.	Terraria	Survival Sandbox	Mix	Yes	Low	Loosely	FAIL 2
16.	CupHead	Run and Gun	Yes	Yes	Low	Closely	FAIL 3
17.	Deep Rock Galactic	Survival FPS	Yes	Yes	Intermediate	Loosely	BACKUP 2
18.	Knights and Bikes	Co-op action adventure	Yes	Yes	Low	Closely	FAIL 4
19.	Sea of Thieves	Action Adventure	Yes	Yes	Low	Loosely	FAIL 5
20.	Borderland 3	Role playing FPS	Yes	Yes	Low	Loosely	FAIL 6
21.	Castle Crasher	2D scrolling hack and slash	Yes	Yes	Low	Closely	FAIL 7
22.	Human fall flat	Puzzle Platformer	Yes	Yes	Low/Intermediate	Closely	BACKUP 3
23.	Unravel 2	Adventure Puzzle Platformer	Yes	Yes	Low/Intermediate	Closely	BACKUP 4
24.	Among us	Social Deduction	Mix	Yes	Low	Closely	FAIL 10
25.	No man sky	Action Adventure Survival	Yes	Yes	Intermediate	Loosely	BACKUP 5

26.	Stardew Valley	Simulation role playing	Yes	Yes	Low	loosely	FAIL 11
27.	Unrailed	Rail-road construction simulation	Yes	Yes	Intermediate	Closely	PASS 13
28.	Biped	Puzzle Platformer	Yes	Yes	High	Closely	PASS 14
29.	Shift Happens	Puzzle Platformer	Yes	Yes	High	Closely	Pass 15
30.	DYO	2D puzzle Platformer	Yes	Yes	High	Closely	PASS 16
31.	Tick Tock a Tale for Two	Asymmetric	Yes	Yes	High	Closely	PASS 17
32.	Ibb and Obb	2D puzzle platformer	Yes	Yes	Low/Intermediate	Closely	BACKUP 6
33.	Space Food Truck		Yes	Yes	Intermediate/High	Closely	PASS 18
34.	JoggerNauts	Co-op switching	Yes	Yes	Intermediate	Closely	PASS 19
35.	39 Days to Mars	Puzzle adventure	Yes	Yes	High	Closely	PASS 20
36.	Death Squared	Puzzle	Yes	Yes	High	Closely	PASS 21
37.	Catastranauts	Space Action Simulation	Yes	Yes	Intermediate/High	Loosely	PASS22
38.	Conduct Together	Vehicle Simulation	Yes	Yes	Interrmediate/High	Closely	PASS23
39.	Rust	Survival	Yes	Yes	Low/Intermediate	Loosely	BACKUP 7
40.	PlateUp	Cooking Simulation	Yes	Yes	Intermediate	Closely	PASS24
41.	SackBoy a big adventure	Platformer	Mix	Yes	Low	Closely	FAIL 12
42.	Nobody Saves the World	Action role-playing dungeon	Yes	Yes	Low	Closely	FAIL 13
43.	Heave Ho	Platform synchronization	Yes	Yes	High	Closely	PASS 25
44.	Factorio	Construction and management	Mix	Yes	Low/Intermediate	Loosely	BACKUP 8
45.	Risk of Rains	Platformer	Yes	Yes	Low	Closely	FAIL 14
46.	Satisfactory	Factory open world simulation	Yes	Yes	Low/Intermediate	Loosely	BACKUP 9
47.	Scrap Nauts	Base-building	Yes	Yes	Intermediate	Loosely	PASS 26
48.	Starbound	Survival Action Adventure	Yes	Yes	Intermediate	Closely	PASS 27
49.	Eco	Survival	Yes	Yes	Intermediate	Loosely	PASS 28
50.	Pitfall Planet	Puzzle Platformer	Yes	Yes	Intermediate/High	Closely	PASS 29
51.	Operation Tango	Asymmetric	Yes	Yes	High	Closely	PASS 30
52.	Superbugs Awaken	Asymmetric	Yes	Yes	High	Closely	PASS 31
53.	Pico Park	Co-op 2D puzzle	Yes	Yes	High	Closely	PASS 32