

Understanding and Evaluating Cooperative Games

Magy Seif El-Nasr*, Bardia Aghabeigi*, David Milam*, Mona Erfani*, Beth Lameman*,
Hamid Maygoli**, Sang Mah***

*Simon Fraser University
Surrey, BC
(magy, baa17, dma35, mea16)@sfu.ca,
beth@bethaileen.com

**New Media Research and
Education
Vancouver, BC
drmaygoli@gmail.com

***Bardel Entertainment
Vancouver, BC
sang@badel.ca

ABSTRACT

Cooperative design has been an integral part of many games. With the success of games like *Left4Dead*, many game designers and producers are currently exploring the addition of cooperative patterns within their games. Unfortunately, very little research investigated cooperative patterns or methods to evaluate them. In this paper, we present a set of cooperative patterns identified based on analysis of fourteen cooperative games. Additionally, we propose Cooperative Performance Metrics (CPM). To evaluate the use of these CPMs, we ran a study with a total of 60 participants, grouped in 2-3 participants per session. Participants were asked to play four cooperative games (*Rock Band 2*, *Lego Star Wars*, *Kameo*, and *Little Big Planet*). Videos of the play sessions were annotated using the CPMs, which were then mapped to cooperative patterns that caused them. Results, validated through inter-rater agreement, identify several effective cooperative patterns and lessons for future cooperative game designs.

Author Keywords

Game design, cooperative patterns, cooperative game design, user experience, testing, engagement.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design.

INTRODUCTION

Cooperative games encourage participation and collaboration; the goal is not to win as a player but as a team of players. Discovering effective cooperative game patterns is an elusive and important problem [personal communication]. Results of our background questionnaire with 60 6-16 year olds revealed that kids are split when it comes to cooperative games. When asked to choose

between cooperative and competitive games, 55% of them preferred cooperative games, while 77% stated that they would like to play games with both options. The industry realizes this. In the past year alone, several AAA titles, such as *Resident Evil 5* (Capcom, 2008) and *Left4Dead* (Valve, 2008), included an optional cooperative mode.

While cooperative design patterns¹ have been around since the inception of games, very few research studies discussed or documented them. Methods for evaluating them are also in their infancy. Most often, user testing groups within game companies evaluate cooperative games using the same methods used to evaluate single player games [personal communication with Electronic Arts team], which are inappropriate for investigating the cooperative aspect of a game. Therefore, there is a need for (a) understanding current successful cooperative patterns and (b) creating methods to evaluate their effectiveness.

This paper aims to address these issues within the context of cooperative video games by discussing three contributions. First, we present a set of cooperative design patterns developed based on analysis of fourteen cooperative games. These patterns extend previous work and present a comprehensive framework for cooperative game analysis. In addition, we outline a set of Cooperative Performance Metrics (CPMs) used to analyze and evaluate cooperative play. These CPMs were used to analyze data collected through a study of a total of 60 participants grouped in 25 sessions with 2-3 participants/session, playing four cooperative games: *Rock Band 2* (Electronic Arts, 2007), *Lego Star Wars* (LucasArts, 2007), *Kameo: The Elements of Power* (Microsoft Game Studios, 2005), and *Little Big Planet* (Sony, 2008). The aim of the analysis was to investigate connections between the CPMs and the cooperative design patterns discussed in the paper. Results from this study revealed several interesting design lessons for building better cooperative games. We present these results as the second contribution of this paper. The CPMs themselves constitute the third contribution of this paper;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

Copyright 2010 ACM 978-1-60558-929-9/10/04....\$10.00.

¹ We use the word pattern here to mean a specific set of design choices concerning rules or mechanics. This should not be confused with software design patterns.

they were developed and validated to evaluate cooperative play.

The paper is structured as follows. After discussing previous work, we detail the cooperative design patterns we developed based on analysis of fourteen co-op games. In discussing these patterns, we examine the process used to derive them, thus addressing their validity. We then explain the CPMs; in discussing these metrics, we detail the process we used to validate them as instruments for analysis and evaluation of cooperative games. We then outline the study we conducted on four cooperative games. We conclude by discussing the findings and their implications, as well as future research.

PREVIOUS WORK

Related works fall within two areas: video games evaluation methods, and cooperative design.

Methods for Game User Studies

Usability testing is an integral part of any software development process. Methodologies of usability and user testing have been addressed in many Human Computer Interaction works [1]. Game evaluation methods integrate many of these HCI methods, but extend them to include testing for playability and engagement [2, 3]. The game industry realizes the importance of developing and conducting game evaluations; this is evident by the formation of groups such as Microsoft's User Experience group, Sony's usability and playtest, Ubisoft's playtest, and Eidos' user-research groups, and the emergence of several user research companies, such as Emsense and XEODesign.

Further, Microsoft developed an online tracking system called TRUE to collect and visualize gameplay telemetry data and synchronize them with attitudinal and observational data [4]. This enabled them to "detect issues and understand root causes in the same way usability testing does [4]." They validated their system using two games: *Halo 2* and *Shadowrun*. In addition, Dracken et al. in cooperation with Eidos interactive developed a set of metrics to track user behavior in *Tomb Raider*. They used Geographic Information Systems to visualize spatial gameplay metrics, developing the Heat Map, which enabled them to detect level design problems, such as places where a lot of players died [5, 6]. Similar methods are used by Valve [7], Bungie and Electronic Arts.

A few studies concentrated on defining methods for evaluating engagement or enjoyment in games. Sweetser and Wyeth developed a model called GameFlow [8] based on the Flow theory [9]. The model consisted of a set of qualitative criteria for measuring eight specific elements of a game: concentration, challenge, skills, control, clear goals, feedback, immersion, and social. They validated the model by evaluating two commercial games, and comparing their results to that of expert reviews. Yannakakis and Hallam [10] explored the development of a

quantitative experimental model specifically targeting simple arcade and augmented reality games. They concentrated on challenge as a main aspect of engagement.

In addition, Lazarro et al. [11] ran a large study with 45 participants with various gaming experience who were asked to play 40+ games of different game genres, including racing, fighting, puzzle solving, and sports. They used observation notes, videotaped interaction, and questionnaires/interviews with friends and family. Based on their study they identified four kinds of fun: (1) hard fun: motivated by achievement, (2) Easy Fun: motivated by exploration, (3) Altered states: motivated by visceral rewards, and (4) Social: motivated by competitive or cooperative play or just to be with friends. Even though the sample was small and included many games, their work contributed data showing variations in play styles and motivations.

An alternative approach to playtesting and usability studies is the use game heuristics evaluation. This method is based on the usability heuristic technique developed by Nielsen [12]. Game heuristic evaluation is accomplished through a systematic inspection of a game using a set of heuristics or guidelines. The technique provides a very cheap and easy to administer testing method, which became very popular within software companies. Several researchers [13-15] worked on developing a set of game design heuristics, expanding on the user interface heuristics develop by Nielson [12]. However, heuristics cannot completely replace playtesting and usability studies, as it does not provide attitudinal, behavior, or play data from actual users.

While previous works present excellent research that addressed the evaluation of games, the measurement and evaluation of cooperative games is still an untapped area. The only work we found targeting this area was Pinelle et al.'s work [16] on heuristics for evaluating networked multiplayer games. While their work and criteria is close to ours, they focused on heuristics based techniques and developed methods to evaluate networked games rather than games where participants share the same physical space. In this paper, we specifically propose a set of validated Cooperative Performance Metrics CPMs for analyzing and evaluating cooperative play occurring over the network or within the same space.

Cooperative Game Design Patterns

Some researchers analyzed a set of cooperative games to develop cooperative game design patterns. Zagal et al., for example, explored cooperative patterns within board games [17]. Also, Bjork and Holopainen [18] presented a large number of game design patterns, which included cooperative and social interaction patterns. In addition, Zagal et al. presented an ontology for analyzing game play [19]. Rocha et al. [20] presented a framework of several cooperative game design patterns. In this paper, we follow in the footsteps of these efforts by extending Rocha et al.'s

model [20]. We chose to extend Rocha et al.'s work, because it uses recently published cooperative games, and thus the design patterns and choices are more up to date, while other works published on cooperative patterns used examples that were over five years old. Note that the patterns described by Rocha et al. overlap significantly with those discussed by Zagal and Bjork.

Rocha et al. identified six cooperative game design patterns:

- *Complementarity*: is one of the most commonly used patterns in co-operative games. It implies that players play different character roles to complement each others' activities within the game.
- *Synergies between abilities*: allows one character type to assist or change the abilities of another. For example, in *World of Warcraft* (Blizzard, 2001), a Shadow Priest can cause an enemy to become vulnerable to shadow damage, which also results in an increase in the damage that Warlocks (another character type) can cause.
- *Abilities that can only be used on another player*: an example can be seen in *Team Fortress 2* (Valve, 2007), where Medics can heal other players.
- *Shared goals*: is a pattern used to force players to work together, such as in *World of Warcraft*, where a group of players are given a single quest with a shared goal.
- *Synergies between goals*: is a pattern that forces players to co-operate together through synchronized goals. For example, the achievement system developed for the Pyro and Medic character classes within *Team Fortress 2* gives Pyros the goal of killing three enemies while ubercharged (being made invulnerable by a Medic). The Medic, on the other hand, has a different goal, which is to ubercharge a Pyro while he/she burns enemies.
- *Special rules* denote rules that are used to enforce cooperation within teams. For example, designers can encode rules to denote specific effects to actions within the game when performed on a friendly player. The idea behind these differences is to promote and facilitate cooperation. A good example is the rule in FPS (First Person Shooter) games that prevents damage when players accidentally shoot other players on the same team, known as *Friendly Fire* modes.

IDENTIFIED COOPERATIVE PATTERNS

In parallel with our research on previous work, we conducted a study analyzing new and old cooperative games, specifically exploring their core mechanics and identifying the interaction models behind their co-op game play. Our initial research resulted in a total of 215 PC and video games that had a multiplayer component. However, most of them included competitive rather cooperative patterns. After an initial review, we selected fourteen games for deeper analysis that included cooperative modes; these were: *Left4Dead* (Valve, 2008), *Resident Evil 5* (Capcom, 2009), *Beautiful Katamari* (Namco Bandai), *Kameo: The Elements of Power*, *Lego Star Wars*, *Wall-E* (THQ, 2008),

Cloning Clyde (Microsoft, 2006), *Guitar Hero III* (Activision, 2007), *Harry Potter and the Goblet of Fire* (Electronic Arts, 2005), *Kung Fu Panda* (Activision, 2008), *Little Big Planet*, *Boom Blox* (Electronic Arts, 2008), *Mario Galaxy* (Nintendo, 2007), and *Army of Two* (Electronic Arts, 2008).

The analysis process took two months to complete. During this time, two researchers analyzed each game in detail using game design theory [21-23] and previous work on cooperative game design [17, 19, 20]. They identified distinct design techniques, including resource sharing, controls (user interface), shared goals and puzzles, and reward structures. They also noted visual design characteristics, such as camera settings. They developed a set of design patterns based on this analysis. For validation, the patterns were reviewed by an independent researcher, who has over 10 years of game industry experience. After his approval, we asked a team of two independent researchers to play all the identified games and develop their own cooperative game design patterns. Although we didn't run a Kappa analysis on the rater agreement, we can report a very high agreement, as researchers identified the same patterns, but have used different terms to denote some of them. At the end of this process, researchers met and discussed the patterns; a final set is discussed below.

We differentiate between cooperative games that support cooperative play through sharing a computer or screen vs. patterns designed for online or distributed collaboration. Benford et al. [24] argued that current interfaces and collaborative environments are not designed to support kids playing together on the same computer, but rather assume that they are collaborating with one another where each one is situated at a different computer. This is important as we discuss cooperative patterns, since we identified the same distinction between cooperative games in the market. All fourteen games reviewed in our study were designed for kids to play together through a shared screen. In these situations, camera set up emerged as an important design component.

We identify the following additional patterns:

- *Camera Setting*: there are three design choices for developing a successful camera in a shared screen co-op games—split screen horizontally or vertically, one character in focus, all characters are in focus (the screen doesn't move unless all characters are near each other).
- *Interacting with the same object*: providing interactive objects that can be manipulated by characters' abilities. In *Beautiful Katamari*, players share a ball. Similarly, in *Little Big Planet*, both players can push or grab one object together.
- Similar to shared goals is *Shared Puzzles*: this pattern is a general category for all cooperative design puzzles, also discussed in [18]. This pattern was observed in games such as *Lego Star Wars* and *Little Big Planet*, where both players encounter a shared challenge or obstacle.

- *Shared Characters*: providing a shared NPC (Non-Player Character) equipped with special abilities that players can assume. This pattern can be seen in *Lego Star Wars*, where both players have the ability to assume a special character, but only one can. This enables discussions among players concerning how to share the character.
- *Special characters targeting lone wolf*: this pattern focuses on the design of NPC characters that target players who are working alone. In *Left4Dead*, the Hunter and Smoker are good examples of this pattern.
- *Vocalization*: are patterns that embed automatic vocal expressions on player characters that alert players of different challenging events. It, thus, encourages players to play close together and support each other.
- *Limited resources*: is concerned with providing a limited number of resources, and thus encourages players to share or exchange resources to research the same goal. *Resident Evil 5* uses this technique; many examples of this pattern can be seen in board games [17].

Table 1 shows example patterns from two games.

Game	Significant Design Pattern
Mario Galaxy	<ul style="list-style-type: none"> • <i>Limited resources</i>: the number of stars collected is a shared resource. • <i>Shared Goal</i>: the goal for both players is to gather a certain amount of stars. • <i>Complementarity</i>: the shadow player supports the player controlling Mario.
Resident Evil 5	<ul style="list-style-type: none"> • <i>Camera Setting</i>: split screen in horizontal mode. • <i>Limited resources</i>: sharing ammo. • <i>Abilities that can only be used on another player</i>: healing the other player. • <i>Shared Puzzles</i>: opening locked doors by solving common puzzles, co-op attacks for defeating strong NPCs, co-op jumping for solving platform puzzles.

Table 1. Example patterns for *Mario Galaxy* and *RE5*.

STUDY DESIGN

Given the cooperative patterns discussed above, we ran a study to investigate how players experience cooperative games that embed these patterns. We ran a study with a total of 60 participants: 18 females (average age=9.81), 42 male (average age=10.4) in a total of 25 sessions. Participants were recruited through bulletin boards, special contact lists, schools, and organizations, such as the Boys and Girls Club. We invited participants to come in groups of 2-4 participants: friends or family for a 3 hour play session. As they came in, they signed a consent form and were interviewed. The first interview included questions about their background, playing habits, and previous gaming experience. After this initial interview, we asked them to play four games in 10 minute-sessions. The games were chosen based on our previous analysis and their

popularity given our target age group (8-12). The selected games were *Rock Band 2*, *Lego Star Wars*, *Kameo*, and *Little Big Planet*. We will use the following abbreviations to denote the games: *RB*, *LSW*, *K*, and *LBP*, respectively. After each play session, participants were interviewed individually to gauge their perceptions on their play experience. For further analysis we videotaped all the play sessions front and back as shown in Figure 1.



Figure 1. Screenshots of participants in a session.

In order to analyze the cooperative nature of these games, we defined several metrics: *Cooperative Performance Metrics* (CPMs). These metrics are associated with observable events within a play session, and thus can be used as a basis for video annotation or structured observation of a cooperative play session.

We created these CPMs through an iterative process involving expert and team reviews. The first initial set of metrics was defined based on several play sessions, where researchers played cooperative games and others observed. These metrics were then reviewed and revised by the team of five researchers involved in this study. The metrics were then used to observe and annotate two pilot cooperative play sessions. The metrics were also sent in parallel to three industry game designers working at Electronic Arts and Square Enix. Based on their feedback and the results observed from the two pilot sessions, we revised the metrics. In a meeting conducted with the research team, three with previous game industry experience, we discussed the metrics and approved the final set, which was used to video annotate the 25 play sessions. We later validated the metrics through an inter-rater agreement method discussed below.

The final set of CPMs developed is as follows:

Laughter or excitement together, which we identified as events, where participants:

- laughed at the same time due to a specific game event;
- expressed verbally that they are enjoying the game, looking for utterances, such as “sweet”, “it is a lot of fun”, etc.;
- shook their heads and showed facial nonverbal behaviors that clearly expressed happiness or excitement.

This behavior was coded by labeling each event in the video that led to laughter or excitement based on the criteria above. As different people have different personalities, it is hard to count just one person and neglect the other, and thus we only labeled events where all participants laughed

together, ignoring instances were one laughed without the other(s). We also imposed the constraint that researchers should label events happening in the same space only once per cause.

Another metric that is central to our work is an event that caused participants to *Worked out strategies*. This was identified when participants:

- talked aloud about solving a shared challenge;
- divided a game zone to different parts in order to divide and conquer;
- navigated the world while consulting with each other.

This is important as it refers to cases during gameplay where an obstacle encourages participants to consult with each other and make a local plan to resolve it. For example, in *Lego Star Wars*, there were different platform puzzles that required players to jump over some specific platforms to open the path. This challenge allowed players to consult with one another and make decisions together.

Another related metric is *Helping* each other. This metric corresponds with helping events. These events come in different varieties. For example, we often found that some players help others by leading them through the game, or by pointing to specific buttons. In *Little Big Planet*, we found many tangible instances of this metric, where participants helped one another by pointing to the controller or by handling the controller for the other player. Thus, we define events that signify this metric as events where players:

- talked about controllers, and how one can use the game mechanics;
- told each other the correct way of passing a shared obstacle;
- saved and rescued the other player while he or she was failing;

In our inter-rater agreement experiments we found that researchers can confuse this tactic with the *Worked out strategies* tactic, especially if participants are helping each other. Thus, we imposed the constraint that researchers should label events under the *Helping* CPM when one player is helping the other and not when both are helping each other.

Global Strategies is a metric we created to refer to events where players take different roles during gameplay that complement each others' responsibilities and abilities. A tangible example of this parameter was observed in *Lego Star Wars*, where one player played the role of Jar-Jar (a character with high jumping capabilities) and the other one tried to support Jar-Jar while facing enemies.

One important problem with cooperative games is the gap between skills which causes players to wait for one another. Most of the time this builds frustration, and thus we developed a metric called *Waited for each other* to label events, where one player waits for the other to catch up.

Another related metric is *Got in each others' way*, which is defined as events where one player leads and the other lags

behind, or when one player wants to do an action, x, and the other wants to take a different action, y, and whereby taking these actions they will inevitably interfere or hinder each other's goals.

RESULTS

We used the CPMs to annotate all game play sessions. A total of 3000 minutes of video data were reviewed and annotated (25 sessions front and back videos, totalling 50 60-minute gameplay videos). One researcher took on this task. He went through all videos and labelled each CPM occurrence. For example, when a laughter event as described above is observed, he marked the video and annotated it by labelling the instance as *Laughter and Excitement Together* CPM. In this section, we discuss the totals, averages, standard deviation, 95% confidence intervals for all CPMs per game. We also discuss paired t-tests evaluating statistical significance of the results.

Furthermore, for each CPM label within the video analysis, the researcher identified a cause based on the cooperative design patterns, specifically: complementarity, synergies between abilities, shared goals, synergies between goals, special rules, camera styles, Interacting with the Same Object (ISO), Shared Puzzle (SP), Shared Character (SC), and Miscellaneous (PM). PM is a miscellaneous category that includes animations, cut scenes, or special elements that are specific to one game. For example, the dance animation in *Little Big Planet* caused much laughter. The mapping between CPMs and cooperative patterns were performed through a qualitative interpretive exercise.

Inter-rater	Kappa for Metrics					
	M1	M2	M3	M4	M5	M6
Session 1	0.88	0.67	0.83	0.86	0.78	1
Session 2	1	0.75	0.86	1	0.60	0.83
Average	0.94	0.71	0.84	0.93	0.69	0.91

Table 2. Inter-rater Agreement (M stands for CPM).

Before discussing the results, we will discuss the validation process we performed to evaluate the reliability of the results. First, to establish face validity, patterns and CPMs were developed through an intensive review process as discussed above. To establish scientific validity, we performed a formal validation process. We asked two independent researchers to rate two sessions given the CPMs and the cooperative patterns identified. All researchers were given an introduction to the CPMs and cooperative patterns and were shown an example of how to apply them using a video-taped gameplay session. Afterwards, they were given two videos of play sessions of *Kameo* and *Lego Star Wars* to analyze. We then performed inter-rater agreement and calculated kappa values [25, 26]. Table 2 shows the results of this process. Based on these results, we found that there were almost perfect agreements

for *Laughter and Excitement Together*, *Helping*, *Global Strategies*, and *Got in Each Others' Way* CPMs; we found substantial agreements for *Worked Out Strategies*, and *Waited for Each Other* CPMs. The kappa values presented are sufficient to establish validity of the approach and the results [25, 26].

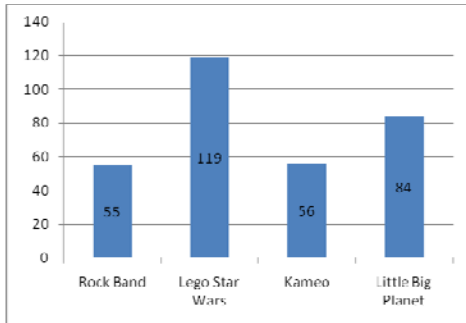


Figure 2. Comparing total number of *Laughter and Excitement Together*.

Laughter and Excitement Together Events

Figure 2 shows totals of events for all sessions labeled as *Laughter and Excitement Together*. Table 3 shows averages per session, standard deviation, and confidence intervals. As it can be seen, *Lego Star Wars* is in the lead with a lot more laughter and excitement events than the rest of the games. *Little Big Planet* follows, then *Kameo* and *Rock Band 2* (same on average). We ran t-tests to check for significance of the differences between the games. T-test results were: RB-LSW (extremely significant, sig = .0001), RB-K (not significant, sig=.9), RB-LBP (significant, sig=.0014), LSW-K (extremely significant, sig=.0003), LSW-LBP (significant, sig=.018), and K-LBP (significant, sig=.009).

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band 2	2.2	1.08	1.77	2.62
LSW	4.7	2.68	3.59	5.74
Kameo	2.24	1.74	1.55	2.92
LBP	3.36	1.87	2.63	3.36

Table 3. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Laughter and Excitement Together*.

Further analysis of the causes of these events reveals that, interestingly, PM is the main cause (shown in figure 3). PM includes a variety of different visual and audio patterns such as character design, character animations, interactive objects, and cut scenes. For example, the falling down animation in *Lego Star Wars* had a great impact on players' excitement. *Little Big Planet's* character designs also had many exciting features such as dancing, shaking hands, etc.. In addition, as the figure shows, shared goals, complementarity, shared puzzles, and shared characters are

important factors, that accounted for 14.1%, 10.2% and 11.4%, respectively.

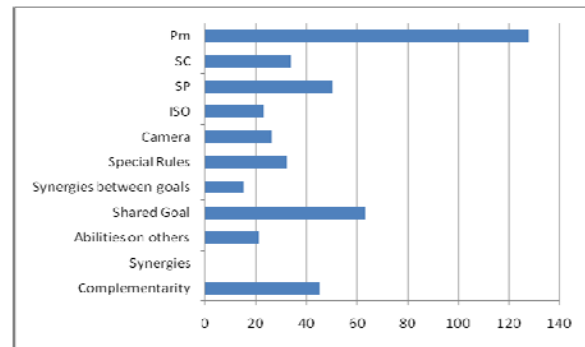


Figure 3. Patterns that caused *Laughter* events.

Worked Out Strategies

Figure 4 shows totals for *Worked Out Strategies* events for all sessions and table 4 shows averages, standard deviation, and confidence intervals. As it can be seen, *Lego Star Wars* is significantly in the lead and *Rock Band 2* is far behind all others with significance. We ran t-test between each pair. T-test results were: RB-LSW (extremely significant, sig = .0001), RB-K (extremely significant, sig=.0001), RB-LBP (extremely significant, sig=.0001), LSW-K (extremely significant, sig=.0001), LSW-LBP (extremely significant, sig=.0001), and K-LBP (not significant, sig=.77).

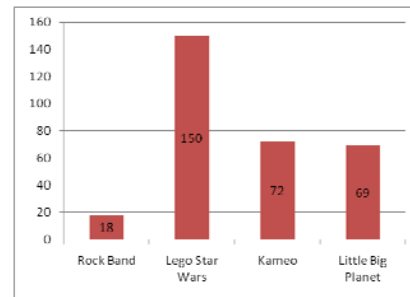


Figure 4. Comparing total number of *Worked Out Strategies*.

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band	.72	0.68	.45	.98
LSW	6.08	2.812	4.95	7.2
Kameo	2.88	1.3	2.37	3.39
LBP	2.76	1.615	2.127	2.76

Table 4. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Worked Out Strategies*.

Figure 5 shows analysis of patterns that caused these *Worked Out Strategies* events. There is a direct impact of shared puzzles and shared goal (60.7%), complementarity (10.8%), shared character (8.1%), and camera pattern (9.1%). As players tried to solve puzzles cooperatively, they talked aloud and made plans. Shared character was

also a cause for these events and was primarily observed in *Lego Star Wars*. Additionally, the complementarity of roles in *Kameo* made this game very challenging, as players switch to different characters to solve puzzles and divide tasks. In one observation, two players worked out their strategies so that one player explored the map while the other fought.

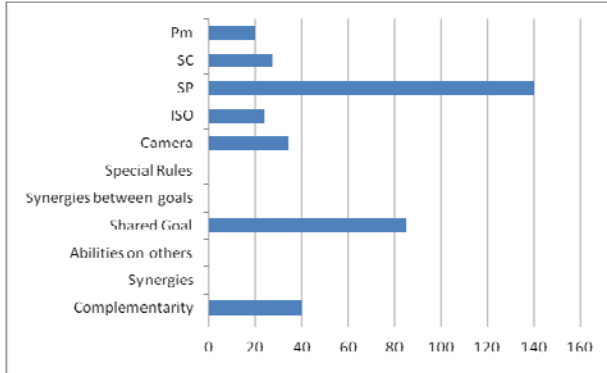


Figure 5. Patterns that caused Worked Out Strategies.

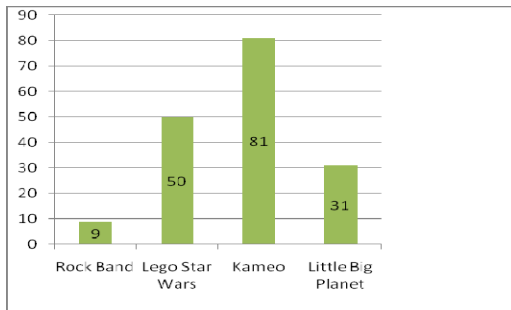


Figure 6. Comparing total number of Helping events.

Helping

Figure 6 shows totals of observed *Helping* events for all session. Table 5 shows averages, standard deviation, and 95% confidence interval per game. The results show that *Kameo* is significantly in the lead here. *Rock Band 2* is last with no overlap with other games. T-test results were: RB-LSW (extremely significant, sig = .0001), RB-K (extremely significant, sig=.0001), RB-LBP (extremely significant, sig=.0007), LSW-K (very significant, sig=.008), LSW-LBP (significant, sig=.034), and K-LBP (extremely significant, sig=.0001).

We deduce from our observation and analysis of gameplay videos that *Kameo* was the most difficult game for players given all the other games. This may be due to the split-screen 3D game. But it was also obvious that many participants had a lot of problems with the controller and the obstacles within the game. This caused them to seek each others' help, and thus may explain the lead of *Kameo*. *Rock Band 2*, on the other hand, is a concentration game that didn't really give players time to help each other.

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band	.36	.7	0.086	.634
LSW	2	1.33	1.43	2.49
Kameo	3.24	1.51	2.65	3.83
LBP	1.24	1.01	.84	1.24

Table 5. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Helping*.

Figure 7 shows a strong relation between *Helping* events and the shared puzzles and goals patterns. These two patterns cover 70% of the *Helping* metric. Also, it is interesting to note synergies between goals as a design pattern accounting for 10% of *Helping* events. *Rock Band 2* was the only game that used this pattern—since players' goals include finishing notes, and the other players' performance has a great impact on group performance.

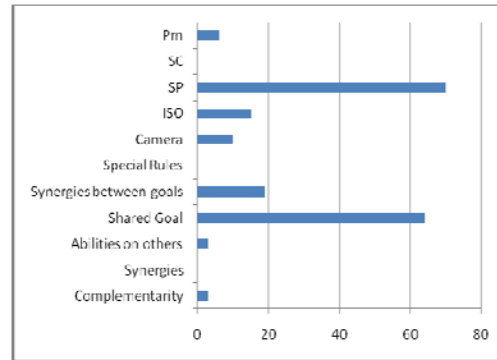


Figure 7. Patterns that caused Helping events.

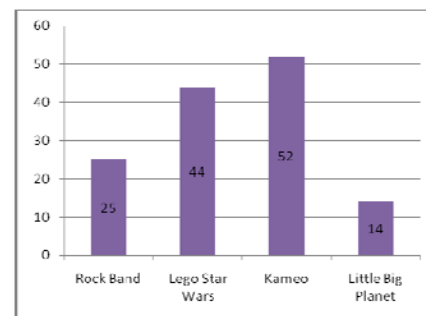


Figure 8. Comparing total number for Global Strategies events.

Global Strategies

Figure 8 shows totals of observed *Global Strategies* events for all sessions; table 6 shows averages, standard deviation, and 95% confidence interval per game. T-test results were: RB-LSW (very significant, sig = .017), RB-K (very significant, sig=.002), RB-LBP (not quite significant, sig=.118), LSW-K (not significant, sig=.246), LSW-LBP (extremely significant, sig=0.0001), and K-LBP (extremely significant, sig=.0001). As it can be seen, there is no

significance between *Kameo* and *Lego Star Wars*, both in the lead. *Rock Band 2* and *Little Big Planet* following. The significant gap between *Kameo* and *Lego Star Wars* on the one hand, and the *Rock Band 2* and *Little Big Planet* on the other, shows that action adventure games support this CPM.

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band	1	1.08	.577	1.42
LSW	1.83	.868	1.486	2.181
Kameo	2.08	1.15	1.63	2.53
LBP	.56	.65	.304	.56

Table 6. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Global Strategies*.

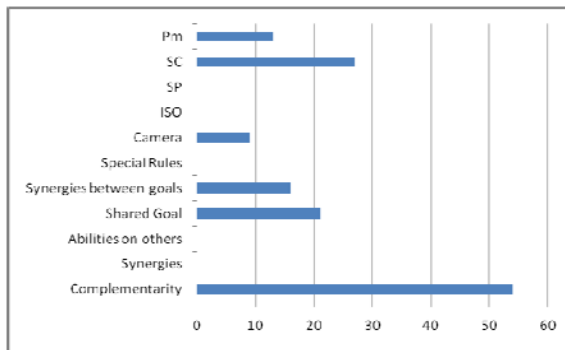


Figure 9. Patterns that caused *Global Strategies*.

Figure 9 shows relations between *Global Strategies* and causes. Complementarity and shared character design patterns account for the majority of these events. Together, they account for 58% of this metric. *Kameo* supports four different characters with different abilities that players switch between dynamically during gameplay. This feature makes it possible for players to assume different roles and develop tactics based on their desired character abilities. Likewise, *Lego Star Wars* uses the shared character pattern named *Jar-Jar*—the player who takes the role of Jar-Jar is responsible for big jumps that solve the platform puzzles in this game, but this character is vulnerable to enemies, and thus the other player has to support him.

Waited for Each Other

Figure 10 shows total events observed for all sessions for the *Waited for Each Other* metric, while table 7 shows averages, standard deviation, and confidence intervals per session. Like with *Global Strategies*, *Lego Star Wars* and *Kameo* are in the lead, overlapping in their confidence interval. Also, *Rock Band 2* and *Little Big Planet* follow with little overlap in their confidence intervals. T-test results were: RB-LSW (extremely significant, sig = .0001), RB-K (extremely significant, sig=.0001), RB-LBP (significant, sig=.031), LSW-K (not significant, sig=.683),

LSW-LBP (very significant, sig=.002), and K-LBP (very significant, sig=.013).

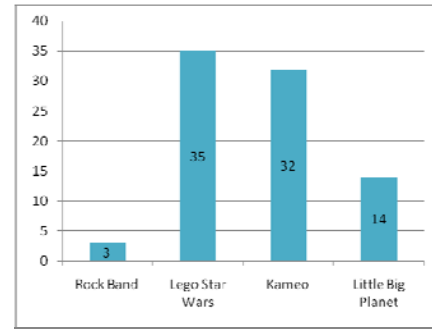


Figure 10. Comparing total numbers for *Wait for Each Other*.

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band	.12	.33	0	.25
LSW	1.4	.977	1.067	1.85
Kameo	1.28	.936	.913	1.647
LBP	.56	.82	.238	.56

Table 7. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Wait for Each Other* CPM.

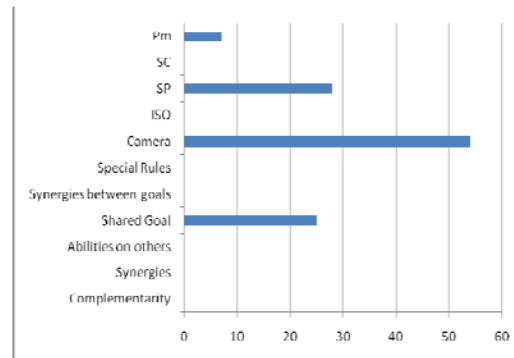


Figure 11. Patterns that caused *Wait for Each Other* events.

Looking at the causes for these events (see Figure 11), it is surprising to see that the camera pattern accounts for 47% of these events. When we take a closer look at the studied games, we see that in *Lego Star Wars*, the camera requires players to wait for each other to proceed. Conversely, *Kameo* has a split screen style, which gives players the freedom to get solve puzzles independently. However, the shared puzzle structures in *Kameo* are designed in such a way that players need to reach the same checkpoints while progressing through the game levels. This caused players to wait for each. It should be noted that *Rock Band 2* has a pausing mechanism that players could use but didn't choose to in any of our sessions.

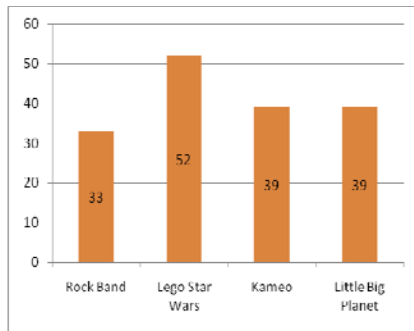


Figure 12. Comparing total number of *Got in Each Others' Way* events.

Game Statistics	Average	Standard Deviation	Lower	Upper
Rock Band	1.32	1.52	.724	1.92
LSW	2.12	1.8	1.4	2.85
Kameo	1.56	1.227	1.07	2.04
LBP	1.56	.96	1.18	1.56

Table 8. Averages, Standard Deviation, 95% confidence Upper and Lower, per game for *Got in Each Others' Way*.

Got in Each Others' Way

Figure 12 shows total of observed events of *Got in Each Others'* for all sessions, and table 8 shows averages, standard deviation, and confidence intervals. As it can be seen, there is overlap between confidence intervals among all games. T-test results were: RB-LSW (significant, sig = .034), RB-K (not significant, sig=.5), RB-LBP (not significant, sig=.5), LSW-K (not significant, sig=.2), LSW-LBP (not significant, sig=.14), and K-LBP (not significant, sig=1). This insignificance may be due to the fact that the CPM was observed for many causes.

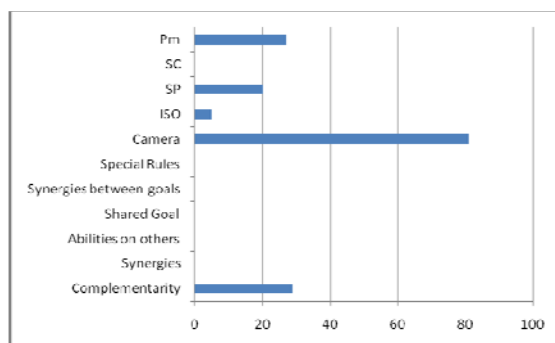


Figure 13. Patterns that caused *Got in Each Others' Way* events.

Camera pattern (50%), complementarity (17%) and shared puzzles (12%) have a great impact on this metric (see Figure 13). The *Lego Stars Wars'* camera depends on players' movements in relation to each other. Thus, if they wanted to move in opposite directions, they will get in each other's way.

CONCLUSION

In conclusion, we present table 9, showing some of the significant cooperative patterns identified based on our results. Specifically, complementarity, shared goals, shared puzzles, and shared objects had a major impact on the identified CPMs. This is evident by the significant results we discussed, specifically in the *Global Strategies* CPM where *Lego Star Wars* and *Kameo* were clearly in the lead due to their use of shared goals, shared puzzles, and complementarity cooperative patterns. In addition, the results suggest that, for the age group we had (6-14), split screen and camera led by the first player caused *Waited for Each Other* and *Got in Each Others' Way* CPMs, which may have a negative impact on the play experience. Thus, designers need to be careful when designing camera settings. Furthermore, analysis of laughter and excitement shows that visual style and animation as well as cut scenes caused much of the *Laughter and Excitement Together* (Figure 3).

Another interesting point to note for cooperative designs is that *Helping* occurred when the game was difficult for players—the number of events observed was significantly higher for *Kameo*, which was rated the most difficult game by our participants. Thus, this CPM is directly tied to difficulty and can be used to tune difficulty of the game.

Game	Pronounced Patterns
Rock Band 2	<ul style="list-style-type: none"> Complementarity Synergies between abilities Abilities on others Shared Goals
Lego Star Wars	<ul style="list-style-type: none"> Complementarity Shared Goal Synergies between goals Camera: all characters are in focus Interacting with same object Shared puzzle and Shared character
Kameo	<ul style="list-style-type: none"> Complementarity Shared Goals and shared Puzzles Interacting with the same object Camera: split screen
Little Big Planet	<ul style="list-style-type: none"> Shared Puzzles Interacting with the same object Abilities on others Camera: led by first player

Table 9. Cooperative patterns leading to positive CPMs.

To summarize, designing effective cooperative patterns is an important area for the game industry, and has a direct impact on the development of educational as well as informal learning games. Developing methods for evaluating or analyzing players' cooperative play is still an untapped research area. In this paper we presented several contributions. First, we proposed several cooperative game design patterns extending previous work. Second, we

proposed a set of Cooperative Performance Metrics (CPMs) used for analysis of the cooperative games. Third, we presented results of a study analyzing the experience of 60 players playing cooperatively in groups of 2-3 four cooperative games: *Rock Band 2*, *Lego Star Wars*, *Kameo*, and *Little Big Planet*. The analysis resulted in valuable design lessons, which form another contribution of this paper. These results were further validated through inter-rater reliability measures. In future research, we will extend this work by running additional experiments with different age groups and game types.

ACKNOWLEDGMENTS

We thank the children and parents who participated in the study. We also wish to thank our funders. The study was funded by MITACS (Mathematics of Information Technology and Complex Systems), a Canadian Network Center of Excellence (NCE), and Bardel Entertainment, a virtual worlds company in Vancouver, British Columbia.

REFERENCES

- [1] A. Sears and J. A. Jacko, *Human-Computer Interaction Fundamentals*: CRC, 2009.
- [2] B. Fulton, "Beyond psychological theory: getting data that improve games," in *Game Developers Conference*, 2002.
- [3] K. Isbister and N. Schaffer, *Game Usability: Advancing the Player Experience*: Morgan Kaufmann, 2008.
- [4] J. P. David, K. Steury, and R. Paygulayan, "A Survey method for assessing perceptions of a game: the consumer playtest in game design," *The international Journal of Computer Game Research*, vol. 5, 2005.
- [5] A. Drachen and A. Canossa, "Analyzing Spatial User Behavior in Computer Games using Geographic Information Systems," in *MindTrek 2009*, 2009.
- [6] A. Tychsen, "Crafting User Experience via Game Metrics Analysis," in *Workshop Research Goals and Strategies for Studying User Experience and Emotion, part of NordiCHI 2008*, Lund, Sweden, 2008.
- [7] L. Nacke, M. Ambinder, A. Canossa, R. Mandryk, and T. Stach, "Game Metrics and Biometrics: The Future of Player Experience Research," in *Future Play*, 2009.
- [8] P. Sweetser and P. Wyeth, "GameFlow: a model for evaluating player enjoyment in games," *Computers in Entertainment (CIE)*, vol. 3, 2005.
- [9] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York: Harper Perennial, 1990.
- [10] N. Yannakakis and J. Hallam, "Capturing Player Enjoyment in Computer Games," in *In Advanced Intelligent Paradigms in Computer Games 2007*, pp. 175-201.
- [11] N. Lazzaro, "Why we play games: Four keys to more emotion without story," XEODesign 2004.
- [12] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the ACM CHI '90*, 1990, pp. 373-80.
- [13] H. Desurvire, M. Caplan, and J. A. Toth, "Using heuristics to evaluate the playability of games," in *CHI*, 2004.
- [14] M. A. Federoff, "Heuristics and usability guidelines for the creation and evaluation of fun in video games." Master's Thesis: Indiana University, 2002.
- [15] N. Schaffer, "Heuristic Evaluation of Games," in *Game Usability*, K. Isbister and N. Schaffer, Eds., 2008.
- [16] D. Pinelle, N. Wong, T. Stach, and C. Gutwin, "Usability Heuristics for Networked Multiplayer Games," in *Supporting Group Work*, 2009.
- [17] J. P. Zagal, J. Rick, and I. Hsi, "Collaborative games: Lessons learned from board games," *Simulation & Gaming*, vol. 37, pp. 24-40, 2006.
- [18] S. Björk and J. Holopainen, *Patterns in Game Design*. California, USA: Charles River Media, 2004.
- [19] J. Zagal, M. Mateas, C. Fernandez-Vara, B. Hochhalter, and N. Lichti, "Towards an Ontological Language for Game Analysis," in *Digital Interactive Games Research Association Conference (DiGRA 2005)*, 2005.
- [20] J. B. Rocha, S. Mascarenhas, and R. Prada, "Game Mechanics for Cooperative Games," in *ZDN Digital Game*, 2008.
- [21] R. Hunicke, M. L. Blanc, and R. Zubek, "MDA Framework for Game Design," in *AAAI 2004: Game AI Workshop proceedings*, San Jose, CA, 2004.
- [22] E. Adams and A. Rollings, *Fundamentals of Game Design*: Prentice Hall, 2006.
- [23] R. Rouse, *Game Design Theory and Practice*: Wordware Publishing Inc., 2000.
- [24] S. Benford, B. Bederson, K. Åkesson, V. Bayon, A. Druin, P. Hansson, J.-P. Hourcade, R. Ingram, H. Neale, C. O'Malley, K. T. Simsarian, and D. Stanton, "Designing Storytelling Technologies to Encourage Collaboration Between Young Children," in *Human Factors in Computing Systems (CHI 2000)*, 2000.
- [25] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37-46, 1960.
- [26] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, pp. 159-174, 1977.