Computers in Human Behavior xxx (2017) xxx-xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com

ELSEVIER

Full length article

A quantitative method for evaluating the complexity of implementing and performing game features in physically-interactive gamified applications

Christian E. Lopez [a], Conrad S. Tucker [a, b, c, *]

[a] Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA 16802, USA
[b] Engineering Design Technology and Professional Programs, The Pennsylvania State University, University Park, PA 16802, USA
[c] Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 15 October 2016
Received in revised form 17 January 2017
Accepted 18 January 2017
Available online xxx

Keywords:
Gamification
Game features
Active game
User engagement
Behavior change
Human performance

ABSTRACT

Gamification aims to implement game features in non-game contexts, with the goal of increasing the motivation of individuals performing a specific task or set of tasks. The tasks themselves, can focus on cognitive behavior change (e.g., overcoming anxiety) or physical behavior change (e.g., overcoming a shoulder injury). Current gamification methods primarily serve as guidelines and principles for the design of gamified applications. Moreover, these methods often overlook the complexity of actually implementing the game features and do not consider the effects that game features have on individuals' ability to perform a *target task*. A knowledge gap exists in understanding the tradeoffs between the complexity of implementing a game feature and the impact it has on increasing individuals' motivation and performance on a particular task or set of tasks. This paper presents a method for evaluating the complexity of implementing game features and the physical effort required to perform the tasks of the application, with a specific focus on physically-interactive gamified applications. Designers will gain a fundamental understanding of how the implementation of specific game features, contributes toward the objective of the application. A case study is presented that focuses on physically-interactive gamified applications in a virtual environment. Empirical results measuring the effects of game features on participants' performance are presented, which provide evidence in support of the metrics proposed in this study. Knowledge gained from this work will inform designers on how to manage their resources more efficiently and predict possible design issues (e.g., not meeting the objective of the application) while creating gamification applications.

1. Introduction

Gamification is an emerging area of research that is gaining interest across a wide range of domains. The term is defined as "*the use (rather than the extension) of design (rather than game-based technology or other game-related practices) elements (rather than full-fledged games) characteristic for games (rather than play or playfulness) in non-game contexts (regardless of specific usage intentions, contexts, or media of implementation)*" (Deterding, Dixon, Khaled, & Nacke, 2011, p. 10). Primarily, a gamified application implements game features with the objective of increasing individuals' motivation towards a target task or set of tasks. Deterding et al. define game features as the "*design elements that can be found in most (but not necessarily all) games*" (Deterding et al., 2011, p. 10).

Gamification researchers have started exploring the benefits of physically-interactive applications. Physically-interactive applications require individuals to use full body motions (e.g., jump, move side to side, bend) to interact with the application. The purpose of these applications is to motivate individuals to perform a *target task*, with the goal of meeting a certain objective(s). The objective can vary depending on the context and the designers' intentions. In the educational context, physically-interactive learning environments have been shown to improve students' motivation and learning (Yang, Chen, & Chang, 2010). Similarly, in the health and wellness context, physically-interactive gamified applications have been shown to improve the physical health of individuals (Biddiss & Irwin, 2010; McCallum, 2012; Read & Shorter, 2011).

Several motivational models and theories indicate that there exists a relationship between individuals' motivational levels, and their ability to perform a task (Csikszentmihalyi, 1990; Fogg, 2009; Oinas-Kukkonen & Harjumaa, 2008; Ryan & Deci, 2000). Several factors can affect the simplicity of a task and the ability of individuals to perform it. The physical effort required to perform a task is among these factors (Fogg, 2009). In physically-interactive gamified applications, the addition of a game feature might affect the performance of individuals by adding *indirect tasks* that are not aligned with the objective of the application. Therefore, gamified applications can have a task or set of tasks that are directly aligned with the objective of the application (*target tasks*), or that are not directly aligned (*indirect tasks*). Fig. 1 shows a representation of how these elements interact in a physically-interactive gamified application. Fig. 1 also illustrates how a gamified application (e.g., Application A) implements differ-

* Corresponding author. 213 N Hammond Building, State College, PA 16803, USA.
Email address: ctucker4@psu.edu (C.S. Tucker)

http://dx.doi.org/10.1016/j.chb.2017.01.036
0747-5632/

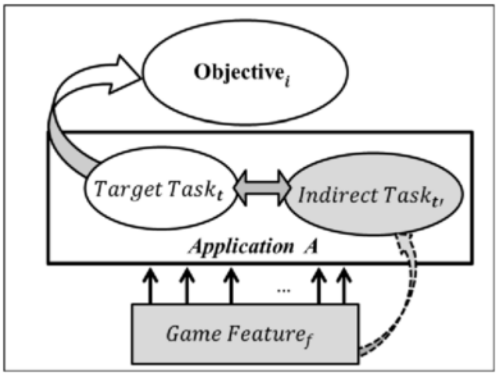

Fig. 1. Physically-interactive gamified application representation.

ent game features (f from the set of game features: {**F**}), with the objective of motivating individuals to perform a *target task t* (*t* from the set of *target tasks*:{**T**}, where $t \geq 1$). This *target task* is aligned with an objective i ($i \geq 1$). This means that by performing this *target task*, the individual will meet the objective of the application. However, individuals' performance on a *target task* might be mediated by the addition of an *indirect task t'* (*t'* from the set of *indirect tasks*: {**T'**}, where $t' \geq 0$). This *indirect task* may unintentionally result from the implementation of a game feature. Therefore, the addition of a game feature may affect the performance of an individual attempting to complete a *target task*. Moreover, the effects of certain game features on individuals' performance may be mediated by other features that are implemented in the same application. In the gamification literature, studies that have implemented numerous game features have only reported positive results for some of the game features employed (Hamari, Koivisto, & Sarsa, 2014). Furthermore, few studies have explored the relationship between the physical effort required to perform the *target tasks* and the *indirect tasks* of gamified applications. Therefore, the effects of the additional effort required to perform an *indirect task* and the impact that game features have on individuals' performance are still unexplored. A better understanding of these relationships and their effects on individuals' performance has the potential to improve the design process of gamified applications.

Even though several studies have shown positive results in individuals' performance by introducing game features into their applications, a limited amount of research has analyzed the complexity of implementing the game features (O'Donovan, Gain, & Marais, 2013). Moreover, the wide availability of game features makes it harder for designers to select and incorporate them into a single application. Consequently, designing and implementing the right set of game features requires designers to spend a significant amount of time and resources (Bharathi, Singh, Tucker, & Nembhard, 2016). While several theories of motivation have been applied to guide the process of designing gamified applications, they do not provide a systematic method that quantifies neither the complexity of implementing the game features (Oinas-Kukkonen & Harjumaa, 2008) nor individuals' performance (Pedreira, Garcia, Brisaboa, & Piattini, 2015). A systematic method that evaluates game features will enable designers to manage their time and resources more efficiently, thus optimizing and reducing the cost of designing gamified applications.

Several studies have shown the positive results of implementing gamified applications in a wide range of domains (e.g., education, health and wellness, and marketing). However, researchers agree that there is a need for more empirical studies that quantify the benefits of gamification (Dicheva, Dichev, Agre, & Angelova, 2015; Hamari et al., 2014; Lucassen & Jansen, 2014; McCallum, 2012; Pedreira et al., 2015). Furthermore, a better understanding of the complexity of implementing game features, their effects on individuals' performance, and their relationship with the physical effort required to perform the *target tasks* of gamified applications is needed. As McCallum states " *there is no magic formula for developing successful* [gamified application]" (McCallum, 2012, p. 92). Hence, the objective of this paper is to present a systematic method that will help designers in the evaluation and selection of game features. The method will quantitatively explore the relationship between the physical efforts required to perform *target tasks* and *indirect tasks* in physically-interactive gamified applications. Additionally, the complexity of implementing game features and their effects on individuals' performance will be analyzed. This information will enable designers to gain a fundamental understanding of how game features contribute to individuals' performance. The remainder of the paper is organized as follows. Section 2 presents the pertinent literature on gamified applications and game features. Section 3 outlines the proposed method. In Section 4, the method is applied to physically-interactive gamified applications in a virtual environment, where quantitative results that support the proposed method are presented. Furthermore, the effects of implementing a set of game features on individuals' performance in gamified applications are discussed. The conclusion and proposal for future works are presented in Section 5.

2. Literature review

2.1. Application of gamification in non-gaming contexts

Even though the concept of implementing game features to improve non-game applications can be traced back to human-computer interaction research of the early '80s (Malone, 1982), the use of the term "*gamification*" is quite recent. Its first documented use was in 2008 (Deterding et al., 2011). However, it was not until after 2010 that the term "*gamification*" experienced widespread adoption (Bharathi et al., 2016; Dicheva et al., 2015; Hamari et al., 2014; Pedreira et al., 2015). Several studies have shown the growing trend in gamification research in fields such as education (Dicheva et al., 2015; Li, Grossman, & Fitzmaurice, 2012; Linehan, Kirman, Lawson, & Chan, 2011; Pedreira et al., 2015), health and wellness (Biddiss & Irwin, 2010; McCallum, 2012; Read & Shorter, 2011), and marketing (Lucassen & Jansen, 2014). Hamari et al. (2014) presented a review of research on gamified applications. Their results suggest that gamification research is becoming a popular research subject. Nonetheless, they also found that numerous studies were descriptive in nature. Hence, there is a need for more empirical evidence in support of the usefulness of gamified applications. Similarly, in their systematic mapping of gamification research in the software engineering fields, Pedreira et al. (2015) concluded that more research is needed that provides empirical results regarding the effects of gamification.

In the educational context, gamified applications are an emergent paradigm that educators are implementing to enhance the learning process (Dicheva et al., 2015). Several studies have reported that gamification improves students' engagement and motivation in a variety of learning activities (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Barata, Gama, Jorge, & Gonçalves, 2016; Caton & Greenhill, 2013). Dicheva et al. (2015) presented a systematic mapping study of gamification research in education. They analyzed different studies based on the design principles, game features, area of application, and academic subjects covered. However, their results also suggest that there is a need for more empirical studies on the effects of implementing gamified applications on improving the learning process.

Some of the studies that provide quantitative evidence in support of gamification are commonly applied to e-learning environments. For example, Li et al. (2012) presented a gamified tutorial system for AutoCAD users. The participants that used the application reported higher levels of engagement and joy. In their study, participants who used the gamified application had faster completion times in a set of testing tasks, when compared to the control group. However, the authors attribute this improvement to the Content Unlocking feature present in the application, which required participants to repeat some segments of the tutorial. Therefore, the overall learning time spent was substantially different between groups. Consequently, their results should be considered with a degree of caution (Li et al., 2012). Similarly, the results by O'Donovan et al. (2013) suggest that with the implementation of well-designed game features, they were able to increase the material reviewed, class participation, and lecture attendance of students. Additionally, they considered the time and monetary cost of implementing this new gamified application. They concluded that the positive results of gamified applications must be balanced against the cost required to successfully implement it.

Several models and theories of motivation indicate that there exists a relationship between the ability of individuals to perform a task and their motivational levels (Csikszentmihalyi, 1990; Fogg, 2009; Oinas-Kukkonen & Harjumaa, 2008; Ryan & Deci, 2000). The Fogg's Behavior Model analyzes the elements relevant to gamification that motivate individuals to perform a task. The model states that for an individual to perform a task, "*he or she must (i) be sufficiently motivated, (ii) have the ability to perform the task(s), and (iii) be triggered to perform it*" (Fogg, 2009, p. 40). Studies, such as the one presented by Denny (2013), provide evidence in support of these models. Denny (2013) presented an application in which an *Achievement* feature was implemented in an online repository of multiple-choice questions. In this application, students were responsible for generating and moderating the learning process and providing peer feedback. The results of the study suggest that the *Achievement* feature had a positive effect on the number of answers submitted and the number of days for which students were active, compared to the control group. However, the game feature did not have any significant effect on the number of questions generated by the students. The author attributed these results to the greater effort and time required to generate questions, compared to the effort and time required to answer them (Denny, 2013). These results are consistent with Fogg's Behavior Model, since if the simplicity of a *target task* or the ability of individuals to perform it decreases, their motivational levels need to increase in order to maintain performance (Fogg, 2009). Several factors can affect the simplicity of a task and the ability of individuals to perform it. The physical effort required to perform a task is among these factors (Fogg, 2009). In physically-interactive gamified applications, this factor can play an important role. Yang et al. (2010) presented a physically-interactive learning environment application that combined video capture and virtual reality technology. Moreover, the students interacted with the application using full body motion. The results suggest that students' long-term learning and motivation improved with the use of the gamified application. However, the authors did not analyze the physical effort required to perform the *target tasks*, nor the *indirect tasks* related to the implementation of the game features. Additionally, the complexity of implementing neither the system nor the individual game features was considered.

In the health and wellness context, there are typically three types of applications based on their objective: (i) ones that aim to improve the physical health, (ii) ones that aim to improve cognitive health and (iii) ones that aim to improve social and emotional well-being (e.g., Wii Fit, Brain Age, and Kinect Sports). Some studies have shown that gamified applications help individuals improve compliance levels and their quality of life (Jibb et al., 2012; Rose, Koenig, & Wiesbauer, 2013; Stinson et al., 2013). Moreover, a growing body of health and wellness research is focusing on physically-interactive gamified applications, such as Active Games, with the objective of improving individuals' physical health (Sell, Lillie, & Taylor, 2008). Active Games are physically-interactive gamified applications that require individuals to apply full body motions (e.g., running, dancing, jumping) to perform *a target task* or set of tasks, such as group exercises, virtual sports or other physical activities (Mears & Hansen, 2009). Biddiss and Irwin (2010) presented a comprehensive literature review of studies that implemented physically-interactive gamified applications to encourage physical activity in children. Their results suggest that these applications enable light to moderate physical activity. However, the papers reviewed did not consider the effects of the different game features on the activity level of children. Similarly, Gao and Chen (2014) presented a literature review of physically-interactive gamified application research with the objective of preventing obesity in children. Their results suggest that these applications are desirable tools for promoting physical activity in children. However, they highlight that future research should study the effect of different game features. Brauner, Calero Valdez, Schroeder, and Ziefle (2013) presented a physically-interactive gamified application aimed at increasing physical fitness and creating health awareness for elderly people. Their results suggest that the application improved the physical health of participants. They analyzed factors such as age, gender, and personality traits but did not consider the game features implemented. Even though the field of gamification in health and wellness has its unique challenges, such as those related to the ethics of clinical trials and their approval process, researchers in the field agree that gamification will be part of the health and wellness future (McCallum, 2012).

The majority of research highlights the potential benefits of correctly implementing well-designed game features for engaging individuals to perform a certain *target task*. Nonetheless, there is a need for more empirical research on the effectiveness of gamification (Dicheva et al., 2015; Hamari et al., 2014; McCallum, 2012; Pedreira et al., 2015). Furthermore, the positive outcomes from the implementation of game features must be balanced against the cost required to successfully implement them (O'Donovan et al., 2013). However, the studies that have shown positive results toward gamification do not take into consideration the complexity of implementing the game features into their applications. There is a fundamental knowledge gap in terms of the tradeoffs that exists between the complexity of implementing game features, and the resulting impact they have on increasing individuals' performance and motivation. Furthermore, in the literature of physically-interactive gamified applications, few studies have explored the relationship between the physical efforts required to perform the tasks of an application. The objective of this paper is to bridge this knowledge gap by presenting a method that quantitatively evaluates the complexity of implementing game features and the physical effort required to perform *indirect tasks* related to the game features. Knowledge gained from this work will inform designers on how to manage their creative resources more efficiently and provide empirical results on the effects of gamified applications on individuals' performance.

2.2. Game design features

Gamification is the implementation of game features in non-game contexts with the objective of increasing individuals' motivation towards a *target task* (Deterding et al., 2011). Deterding et al. (2011)

define game features as "*the design elements that can be found in most (but not necessarily all) games*" (Deterding et al., 2011, p. 10). Werbach and Hunter (2012, p. 82), proposed a classification of game features as:

- "***Dynamics***- *are the big-picture, aspects of the gamified systems that you have to consider and manage but which you never directly enter into the game*"
- "***Mechanics***- *are the basic processes that drive the action forward and generate player engagement*"
- "***Components***- *are the specific instantiations of mechanics or dynamics*"

Bharathi et al. (2016) applied a Sequential Minimal Optimization algorithm with the objective of identifying the set of game features common among successful and unsuccessful mobile task-driven applications. They analyzed 60 different applications and used their ranking score on the Google Play™ store to measure the success of the applications. They showed that with the use of 24 different game features, they were able to predict if an application was successful or not, based on their ranking score. Table 1, obtained from Bharathi et al. (2016), shows the individual game features and their definitions arranged in decreasing order of their coefficient obtained from the Sequential Minimal Optimization algorithm. Their results suggested that there is a set of game features that are shared by successful applications (ranking 1 to 15) and another set of game features shared by unsuccessful applications (ranking 16 to 24, gray background). However, the authors only evaluated if the game features were present or not. Hence, how the game features were implemented was not considered in the analysis. This factor may affect the success or failure of

Table 1
Ranked game features and definition from Bharathi et al. (2016 p. 364).

Ranking	Game features	Definitions
1	Points	*"Numerical representation of game progression"*
2	Avatars	*"Visual representations of players' characters"*
3	Challenges	*"Puzzles or other tasks that require effort to solve"*
4	Virtual Goods	*"Game assets with perceived or real money value"*
5	Competition	*"Getting players to compete against one another"*
6	Boss Fights	*"Especially hard challenges at the culmination of a level"*
7	Teams	*"Defined group of players working towards a common goal"*
8	Leaderboards	*"Visual displays of players progression and achievements"*
9	Gifting	*"Gives an opportunity to gift things such as lives/ points to other players"*
10	Content unlocking	*"Unlocks new levels/new features when players reach specific objectives"*
11	Transactions	*"Buying, selling or trading with other human players or automated players"*
12	Turns	*"Sequential participation by alternating players"*
13	Quests	*"Predefined challenges with objectives and rewards"*
14	Cooperation	*"Getting players to work together to achieve a shared goal"*
15	Feedback	*"Information about how the player is doing"*
16	Badges	*"Visual representations of achievements"*
17	Win states	*"The state that defines winning the game"*
18	Levels	*"Defined steps in player progression"*
19	Rewards	*"Some benefits that go together for some action or achievement in the game"*
20	Collections	*"Set of items or badges to accumulate"*
21	Resource acquisition	*"Obtaining useful or collectible item"*
22	Chance	*"Involvement of luck from a random mechanism"*
23	Social graph	*"Ability to track progress of friend and enables interaction"*
24	Achievements	*"A form of reward attached to performing specific actions"*

an application in motivating individuals to perform a *target task*. Nonetheless, the importance of their work is that it can be used as a starting point for the game feature selection process.

Moreover, the results presented by Dicheva et al. (2015) shows that the most popular game features used in educational applications were Points, Badges, and Leaderboards; followed by Levels, Virtual Goods, and Avatars. Similarly, the results presented by Hamari et al. (2014) shows that most of the gamified applications reviewed, implemented Points, Badges, and Leaderboards. However, they also revealed that the majority of studies that implemented numerous game features reported positive results for only some of the game features employed. Similarly, the results presented by Pedreira et al. (2015) suggests that most of the studies analyzed, implemented game features such as Points, Levels, and Badges. In more than half of the studies they analyzed, the only game features applied were Points or Badges. It can be seen that multiple studies have shown the use of a limited and non-homogenous set of game features (Dicheva et al., 2015; Hamari et al., 2014; Pedreira et al., 2015). While the game features of Leaderboard, Virtual Goods, and Avatar were present in some of the papers related to gamification in the educational context reviewed by Dicheva et al. (2015), they were not present in any of the papers analyzed by Pedreira et al. (2015). Similarly, the game features of Quest and Social Graph were present in some of the papers related to gamification in the software engineering field reviewed by Pedreira et al. (2015), but they were not present in any of the papers analyzed by Dicheva et al. (2015). These results can be understood by the definition of game features found in the literature (see Table 1). From these definitions, it is clear that some game features are not mutually exclusive and others would be impractical to apply in certain contexts. For example, the game features of Competitions ("*getting players to compete against one another*") and Cooperation ("*getting players to work together to achieve a shared goal*") would be impractical to apply without the option of multiple players. Moreover, the game features of Rewards ("*some benefits that go together for some action or achievement in the game*") and Achievements ("*a form of reward attached to performing specific action*") are not mutually exclusive (see Table 1 for more definitions).

As Dicheva et al. (2015) stated: "*gamification has the potential to improve learning if it is well designed and used correctly*" (Dicheva et al., 2015, p. 83). However, the wide availability of game features and their interdependence makes it difficult for designers to correctly select and incorporate them (Bharathi et al., 2016). Furthermore, no studies have demonstrated which game features have the greatest impact on increasing individuals' motivation (Pedreira et al., 2015). Hence, designers of gamified applications currently do not have a systematic method for developing successful gamified applications (McCallum, 2012). Even though the results presented by Bharathi et al. (2016) suggests that some of the game features correlate to successful applications, no empirical evidence suggest that this set of game features will improve individuals' motivation to perform a *target task*. Furthermore, even though several theories have been proposed to guide the design process of gamified applications, they do not provide any quantitative nor systematic method (Oinas-Kukkonen & Harjumaa, 2008). Therefore, the objective of this paper is to present a systematic method that will guide designers in the evaluation and selection of game features. The application of this method will allow designers to gain a fundamental understanding of which features are worth exploring. Furthermore, quantitative results of the effects of different game features on individuals' performance are presented. This information will enable designers to gain a fundamental understanding of how certain game features contribute to the *target tasks* of a gamified application.

3. Method

In this section, a method is presented that aims to quantitatively explore the complexity of implementing game features and the physical effort required to perform tasks in physically-interactive gamified applications. This method will guide designers towards the systematic discovery of the game features that are worth exploring and incorporating into their applications. Fig. 2 presents the outline of the proposed Feature Discovery Method that includes the Initial Features Selection (3.1), Game Feature Complexity Analysis (3.2), and Key Features Discovery (3.3). The Model Validation (3.4) section outlines the procedure that will help validate the metrics and steps proposed in the Feature Discovery Method. In this section, the research hypotheses and their theoretical arguments are presented.

3.1. *Initial Feature Selection*

To help designers understand the set of game features that are practical to apply in their gamified applications, an Initial Feature Selection algorithm is proposed. The algorithm, illustrated in Fig. 3, can be used as a starting point for the design of gamified applications. The algorithm shows the relationships between game features and the possible constraints imposed by the information technology systems used. Each of the if-else rules and the clustering of non-mutually exclusive game features presented in the algorithm, are based on their definitions found in the literature (see Table 1, section 2.2). The algorithm uses three binary variables as inputs. These variables are aimed to capture the presence or absence of constraints imposed by the information technology system. These constraints could prevent designers from successfully implementing a game feature. The variables are defined as follows:

1) $\alpha = 1$, if the individuals will be allowed to play more than once; otherwise $\alpha = 0$.
2) $\beta = 1$, if the application will be connected to a network that will enable it to interact with other applications; otherwise $\beta = 0$.
3) $\lambda = 1$, if the individuals will be allowed to interact with multiple players or an Artificial Intelligent system; otherwise $\lambda = 0$.

After defining these variables, designers can use them as input for the algorithm shown in Fig. 3. As output, designers will obtain a subset of game features that will be practical to implement in their gamified application, based on the constraints imposed by the information technology systems that will be used as a communication channel. For example, if the only constraint of the information technology system is that it will not allow an individual to interact with another player or Artificial Intelligent system ($\lambda = 0$, $\alpha = 1$, and $\beta = 1$), then the game features of Teams, Competition, Cooperation, Leaderboards, Transaction, Gifting, Social Graphs, or Turns will not be practical to implement since by definition, they require multiple players (e.g., Teams: "*defined group of players working towards a common goal*", Competitions: "*getting players to compete against one another*", Cooperation: "*getting players to work together to achieve a shared goal*", see Table 1, section 2.2 for other game feature definitions). After the initial game features are selected, the complexity of implementing the game features needs to be assessed. Subsequently, the physical effort required to perform any *indirect task* related to the game features selected needs to be analyzed. The Game Feature Complexity Analysis is presented in section 3.2.

3.2. *Game features complexity analysis*

3.2.1. *Game feature implementation complexity*

The goal of measuring the implementation complexity of the different game features selected is for designers to gain a better understanding of the resources required to implement them. This information will allow them to manage their resources more efficiently. The method proposes the use of an Implementation Complexity metric to measure the complexity of implementing a game feature. This metric is based on the Information Flow metric proposed by Henry and Selig (1990). The Information Flow and other analogous metrics have been extensively used in the software literature to measure the complexity of implementing, maintaining, and/or understanding different applications (Jabangwe, Börstler, Šmite, & Wohlin, 2015; Mens, 2016; Phukan, Kalava, & Prabhu, 2005). The foundation behind Information Flow is that the complexity of an application is a function of (i) the internal complexity of its features ("*intra-module complexity*") and (ii) the complexity of the feature interactions ("*inter-module complexity*"). The code size metric of Lines of Code is often used to measure the intra-module complexity. The predominant definition for Lines of Code is "*a line of program text that is not a comment or blank line, regardless of the number of statements or fragments of statements on the line*" (Aggarwal & Singh, 2005, p. 131). However, this metric is highly dependent on the designer coding capabilities and the language used. Additionally, it cannot be applied to visual languages because the notion of Lines of Code is not meaningful (Phukan et al., 2005). On the other hand, the inter-module complexity, which considers the overall exchange of information between features, does not depend on designers' coding abilities nor the language used (Phukan et al., 2005). Therefore, the Implementation Complexity metric proposed in this method will only consider the inter-module complexity of a feature. Henry and Selig (1990) defined a con-

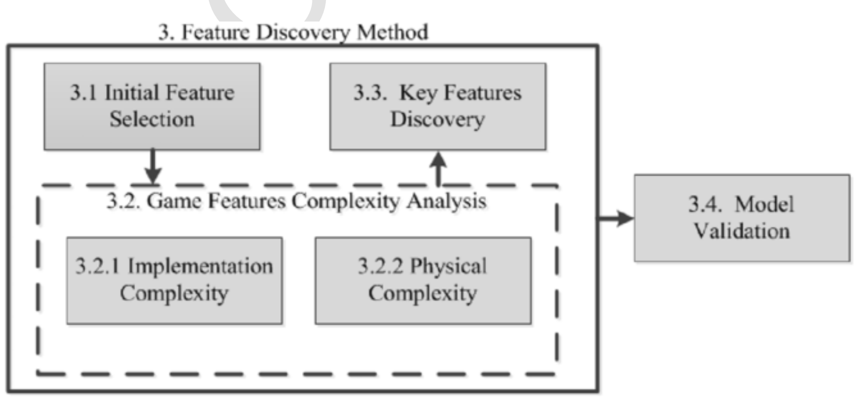

Fig. 2. Method diagram.

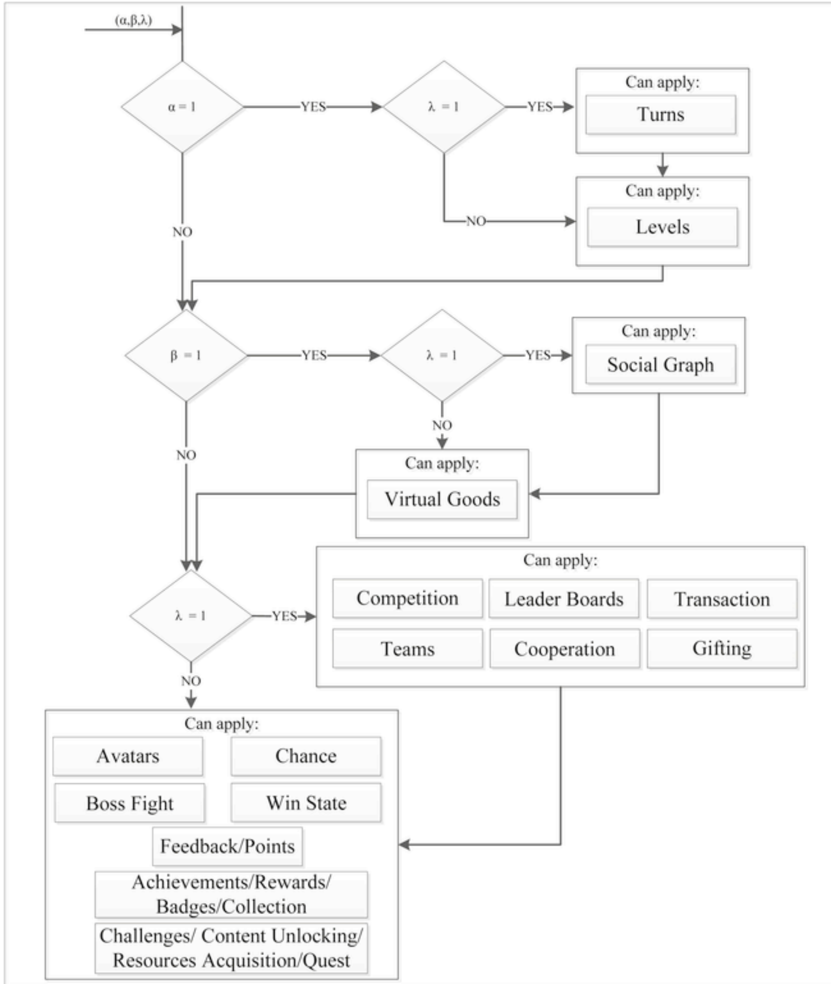

Fig. 3. Initial feature selection algorithm.

nection between two features (or a feature and a global data structure) to exist if information is exchanged between them. To calculate the Implementation Complexity metric, designers first need to identify the information flow of the different features. There are four important types of information flow, defined by Henry and Selig (1990) as follows:

1. **Local Direct Flow out (LDF_{out})**: "*If a feature evokes a second feature and passes information to it*"
2. **Local Direct Flow in (LDF_{in})**: "*If the invoked feature returns a result to the caller*"
3. **Global Flow out (GF_{out}): "***If a feature updates information from a global data structure*"
4. **Global Flow in (GF_{in}): "***If a feature evokes a global data structure to retrieve information*"

Once the different information flows have been identified, the Implementation Complexity metric of a feature $f\,(\mathrm{IC_f})$ can be calculated as shown in Eq. (1).

$$\boldsymbol{IC_f} = \left(fan-in_f + fan-out_f\right) \quad \text{or } f \in \text{Set of Game Features}\{\mathbf{F}\} \tag{1}$$

Where,

- $fan\text{-}in_f$ is the sum of all Local Direct Flow in (LDF_{in}) and Global Flow in (GF_{in}) for game feature f. Mathematically, this can be express as:

$$\boldsymbol{fan-in_f} = \sum LDF_{in}^{f} + \sum GF_{in}^{f}, \quad for\ \ f\ \varepsilon\ \mathbf{F} \tag{2}$$

- $fan\text{-}out_f$ is the sum of all Local Direct Flow out (LDF_{out}) and Global Flow out (GF_{out}) for game feature f. Mathematically, this can be express as:

$$\boldsymbol{fan-out_f} = \sum LDF_{out}^{f} + \sum GF_{out}^{f}, \quad for\ \ f\ \varepsilon\ \mathbf{F} \tag{3}$$

A feature with a relatively high Implementation Complexity value indicates a possible stress point in the system. This suggests that any modification done to this feature would have the tendency to affect other components of the application, therefore making it challenging for designers to implement (Phukan et al., 2005). A gamified application example is used to show the process of calculating the Implementation Complexity metric. Fig. 4 shows a data flow diagram of an example application that is composed of three game features. In a data flow diagram, the information flow is represented by the arrows

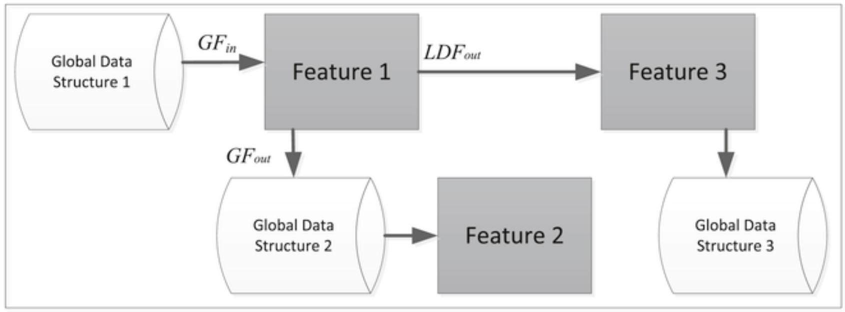

Fig. 4. Data flow diagram of a gamified application example.

and their orientation. The information flows of Feature 1 have been labeled for visualization purposes.

Table 2 shows the calculation of the Implementation Complexity (IC) metric for the gamified application example shown in Fig. 4. The first column of Table 2 shows the feature's number from the example application. The inter-module complexity for the game feature is determined by the total information flows. In columns 2, 3, 5, and 6 of Table 2 the different information flows (LDF_{in}, GF_{in}, LDF_{out}, *and* GF_{out}) of the features are shown. These can be identified by visually inspecting Fig. 4. For illustration proposes, it was assumed that Feature 1 evokes the Global Data Structure 1 to retrieve information, updates information from the Global Data Structure 2, and evokes Feature 2 to pass information to it (see labels on Fig. 4). Therefore, Feature 1 has $LDF_{in} = 0$, $GF_{in} = 1$, $LDF_{out} = 1$, and $GF_{out} = 1$ (see Table 2). The calculated fan-in and fan-out using Eqs. (2) and (3) are shown in columns 4 and 7, respectively. Finally, the calculated Implementation Complexity (IC) values for the game features are shown in column 8. In this example, Feature 1 has a fan-in of 1 and a fan-out of 2. Following Eq. (1), it is clear to show that the Implementation Complexity of Feature 1 is 3 (see Table 2). The second part of the Game Feature Complexity Analysis section relates to the evaluation of the physical effort required to perform the *indirect tasks* resulting from implementing game features. The following section presents the metric proposed in this method to capture the physical effort required to perform the *indirect tasks* of an application.

3.2.2. *Game feature Physical Complexity*

The method employs body movement data to evaluate the physical effort required to accomplish a task. Human skeletal data inferred from individuals' body movement patterns is becoming more adopted by researchers. This is mainly because of the increased availability of motion capture technologies (Behoora & Tucker, 2015; Han, Reily, Hoff, & Zhang, 2016, pp. 1–21). Therefore, researchers can now easily plot three-dimensional coordinates of an individuals' joints with the use of non-wearable infrared sensors (Behoora & Tucker, 2015). The method proposed in this study implements an off the shelf, low-cost infrared sensor (e.g., Microsoft Kinect, or Asus Xtion Live) to acquire participants' skeletal data. The sensors can collect X, Y, and Z distance coordinate data from a reference point, for k joints. However, depending on the sensor used, the number of joints tracked can vary (Behoora & Tucker, 2015). Fig. 5, part A, shows a representation of a human skeletal system with X, Y, and Z distance coordinate data points of the joint representing the right hand. The values of these coordinates measure the relative distance from a reference point, which in this example is located at the vertex of the three axes (shown as a dotted circle). The physical effort required to perform any *indirect task* resulting from the implementation of a game feature will be measured using participants' skeletal data. Similarly, the physical effort required to perform the *target tasks* of a physically-interactive application can be measured using this type of data as well. The method proposes the use of a Physical Complexity metric to evaluate the physical effort required to perform the different tasks of the application (*indirect tasks* and *target tasks*). The Physical Complexity metric of a task t (PC_t) can be calculated as the sum of the Euclidian distance from the joints' positions while at rest (X^{rest}_j, Y^{rest}_j, Z^{rest}_j) to the joints' positions needed to perform the task t (X^t_j, Y^t_j, Z^t_j), of all joints $j = 1$ to k, as shown in Eq. (4). The task t can be from the set of *indirect tasks* $\{\mathbf{T'}\}$ or the set of *target tasks* $\{\mathbf{T}\}$. In this study, an individual is considered to be in resting position when he/she is stand-

Table 2
Implementation Complexity from the gamified application shown in Fig. 4.

Feature No.	LDF_{in}	GF_{in}	Fan-In	LDF_{out}	GF_{out}	Fan-out	IC
1	0	1	1	1	1	2	3
2	0	1	1	0	0	0	1
3	1	0	1	0	1	1	2

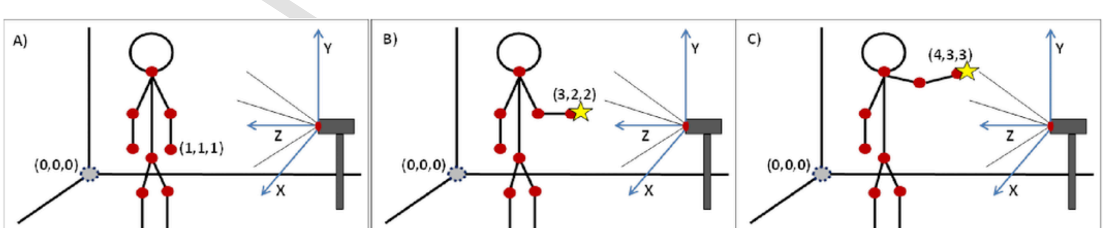

Fig. 5. Representation of a human skeletal joint system.

ing up with his/her arms close to the body (e.g., part A, Fig. 5).

$$PC_t = \sum_{j=1}^{k} \sqrt[2]{\left(X_j^{rest} - X_j^t\right)^2 + \left(Y_j^{rest} - Y_j^t\right)^2 + \left(Z_j^{rest} - Z_j^t\right)^2} \quad \text{for } t \in \text{ set of tasks} : \{\mathrm{T} \text{ or } \mathrm{T}'\}$$

Designers can use sensors to collect human skeletal data from a pilot test of the application. With this data, the position of the participants' joints while at rest (see part A, Fig. 5) and while performing a task (see part B and C, Fig. 5) can be mapped and represented in a three-dimensional coordinate system. The coordinate data of the joints' positions can then be used to measure the Physical Complexity metric, as shown in Eq. (4). Fig. 5, part B and C illustrate a representation of a human skeletal system while performing a task of extending the arm to collect a star. The task shown in part B requires a different level of physical effort to perform than the task in part C. In this example, the coordinates of the node that represent the right-hand side while at rest, as shown in part A Fig. 5, are $X^{rest} = 1m$, $Y^{rest} = 1m$, and $Z^{rest} = 1m$ (for this example, the joint coordinates are given as distance in meters $[m]$ from the reference point X = 0, Y = 0, Z = 0). Additionally, the coordinates of the node that represent the right-hand side while performing the task on part B Fig. 5 (t = B), are $X^B = 3m$, $Y^B = 2m$, and $Z^B = 2m$. Similarly, for part C (t = C), the coordinates are $X^C = 4m$, $Y^C = 3m$, and $Z^C = 3m$. Using Eq. (4) it can be shown that the Physical Complexity metric of the task in part C ($PC_C = 4.1m$) is greater compared to the one in part B ($PC_B = 2.44m$). This suggests that a greater physical effort is needed to perform the task in part C since the location of the star requires individuals to move a greater distance. Moreover, since the Physical Complexity metric will be used to compare the differences in the physical effort required to perform tasks within the same application, it is robust against the distance units used. This is under the assumption that the same sensor is used to collect all the participants' skeletal data.

The accuracy of the Physical Complexity metric on capturing the physical effort required to perform a task will depend on two factors. First, the set of joints involved in the body movements required to perform the task, and secondly, the subset of these joints that are successfully tracked. Therefore, the body movements required to perform the tasks of an application needs to be assessed before selecting a tracking sensor. This is important since the number of joints tracked will vary depending on the sensor used. However, studies have shown that with a subset of joints (see Table 3), activities that require full body movements, such as running, walking, jumping, dancing, playing soccer, and doing Yoga, can be accurately characterized (Chan, Loh, & Rahim, 2016; Shen, Yang, & Liao, 2011; Xiao, Nait-Charif, & Zhang, 2008, pp. 144–153). Therefore, to evaluate the physical effort required to perform the tasks of physically-interactive gamified applications designers need to select a sensor that is capable of tracking at least the joints listed in Table 3. If more joints than the ones listed in Table 3 are tracked, the accuracy of the Physical Complexity metric on capturing the physical effort required to perform a task can only improve (see Mathematical Proof, Appendix A). The Physical Complexity metric will provide designers with valuable information about the physical efforts required to perform the *target tasks* and *indirect tasks* of an application. This information can be used to gain a fundamental understanding of how certain game features affects the individuals' performance. With the Implementation Complexity and the Physical Complexity metrics designers can make a better decision of what features are worth exploring. In the next section, a decision-making process is presented to aid designers of physically-interactive gamified applications in the evaluation and selection of features.

Table 3
Set of body joints used to track full body motion, from Chan et al., 2016.

Body joints			
1	Head	9	Right wrist
2	Neck	10	Left toe
3	Pelvis	11	Left ankle
4	Left wrist	12	Left knee
5	Left elbow	13	Right knee
6	Left shoulder	14	Right ankle
7	Right shoulder	15	Right toe
8	Right elbow		

3.3. Key Features Discovery

The data of the Implementation Complexity and the Physical Complexity metrics will allow designers to systematically discover what features are worth exploring. From an implementation point of view, designers should focus on implementing the game features that are less complex to implement. Additionally, designers need to minimize the physical effort required to perform any *indirect task* related to the implementation of a game feature. Therefore, designers are faced with a minimization problem, in which the objective function is to minimize the Implementation Complexity and the Physical Complexity of a game feature f from the set of game features $\{\mathbf{F}\}$ under consideration. Eq. (5) shows a mathematical representation of this problem that can aid designers in this decision. The metrics are normalized to a range of [0–1] in order to calculate the Discovery metric of a game feature f (θ_f). Therefore, Discovery metric values could range from [0–2].

$$Min\theta_f = \left(\frac{IC_f - IC_{min}}{IC_{max} - IC_{min}}\right) + \left(\frac{PC_{t'} - PC_{min}}{PC_{max} - PC_{min}}\right), \quad f \in \mathbf{F} \tag{5}$$

Where,

- IC_f: The Implementation Complexity metric of the game feature $f \varepsilon$ **F**, as defined in section 3.2.1
- IC_{max}: The maximum value of Implementation Complexity metric from the set **F**
- IC_{min}: The minimum value of Implementation Complexity metric from the set **F**
- $PC_{t'}$:The Physical Complexity metric of the *indirect task t'* that relates to the game feature f, as defined in section 3.2.2
- PC_{max}: The maximum value of Physical Complexity from the set of *indirect tasks* $\mathbf{T'}$
- PC_{min}: The minimum value of Physical Complexity from the set of *indirect tasks* $\mathbf{T'}$
- θ_f: Discovery metric of the game feature $f \varepsilon$ **F**

The Discovery metric values can be ranked in ascending order and used as a criterion to quantify which game features are worth exploring. A low Discovery metric value suggests that a game feature is less complex to implement and that the *indirect task* related to it will require less physical effort to perform in comparison with the other game features under consideration. This new systematic method for evaluating game features in physically-interactive gamified applications provides quantitative information that can help designers in deciding which game features are worth implementing. Knowledge gained from implementing this new method will inform designers on how to manage their creative resources more efficiently and how to optimize the design process.

3.4. Model validation

This section presents the procedure and statistical hypotheses tests that will provide quantitative evidence to validate the metrics proposed in the previous sections. In this study, an alpha value of 0.05 will be used to test the significance of the results. The purpose of physically-interactive gamified applications is to motivate individuals to perform a *target task* or set of tasks with the goal of meeting certain objectives. In these applications, the *target tasks* are designed such that by successfully performing them, individuals will meet the objective of the applications (e.g., *objective*: improve the physical fitness of individuals, *target tasks*: perform jumping jack, pushups, etc.). Due to this relationship, individuals' performance on a *target task* has been used as a proxy for measuring individuals' performance on meeting the objective of the application (e.g., measuring attendance, participation, and quiz grades as a proxy for improved learning, see section 2.1). This method proposes to measure participants' performance on the *target tasks* of an application as a proxy for their performance on meeting the applications' objective. The participants' performance on the *target tasks* is measured as the deviation from the target body position. An Intensity of Mistake metric for a *target task t* (IM_t) is proposed as shown in Eq. (6).

$$IM_t = \sum_{j=1}^{k} \left(|X_j^t - X_j^{pt}| + |Y_j^t - Y_j^{pt}| + |Z_j^t - Z_j^{pt}| \right) * Time_j^t \quad \text{for}$$

Where,

- $X^{pt}{}_j$, $Y^{pt}{}_j$, *and* $Z^{pt}{}_j$: the participants' *j* joint position coordinates while performing a *target task t*, for $t \in \mathbf{T}$, $j \in \mathbf{K}\{1, \ldots, k\}$
- $X^t{}_j$, $Y^t{}_j$, *and* $Z^t{}_j$: the participants' *j* joint position coordinates needed to successfully perform a *target task t*, for $t \in \mathbf{T}$, $j \in \mathbf{K}\{1, \ldots, k\}$
- $Time_j^t$: the total time that the position of a joint *j*, while performing the *target task t*, deviated from the joint position needed to successfully perform it, for $t \in \mathbf{T}$ & $j \in \mathbf{K}\{1, \ldots, k\}$

Based on several motivational model and theories, the Physical Complexity metric proposed in this method should be correlated with participants' performance (see section 2.1). Hence, the Discovery metric, that is a function of the Physical Complexity metric, should be correlated to participants' performance as well. Therefore, the hypotheses that the correlation between the Intensity of Mistake (IM) and Physical Complexity (PC) values for the *indirect tasks* and *target tasks* are statistically significantly different than 0 will be tested. These hypotheses can be express as:

1) $\mathrm{H_o}: \rho_{IM_t, PC_t} = 0$ vs $\mathrm{H_a}: \rho_{IM_t, PC_t} \neq 0, \quad \text{for } t \in Target\ Tasks\ \{\mathbf{T}\}$

2) $\mathrm{H_o}: \rho_{IM_{t'}, PC_{t'}} = 0$ vs $\mathrm{H_a}: \rho_{IM_{t'}, PC_{t'}} \neq 0, \quad \text{for } t' \in Indirect\ Tasks\ \{\mathbf{T'}\}$

Furthermore, to provide empirical evidence to support the results presented by Bharathi et al. (2016), two different applications are used in the case study presented in the next section. Hence, evidence to support that there is a significant difference in performance between participants that interacted with applications that implemented different sets of game features needs to be provided. Therefore, the hypothesis that the mean Intensity of Mistake of individuals that interacted with application A ($\mu_{IM|A}$) is significantly different than the one of individuals that interacted with application B ($\mu_{IM|B}$), will be tested. Given that application A implemented a different set of game features than application B. This hypothesis can be express as:

3) $\mathrm{H_0}: \mu_{IM|A} = \mu_{IM|B}$ vs $\mathrm{H}_a: \mu_{IM|A} \neq \mu_{IM|B}$

4. Case study

This section presents the implementation of the method proposed in section 3 in a set of physically-interactive gamified applications, which only differ in the set of game features used. The purpose of the applications is to motivate individuals to perform certain body motions (e.g., jump, bend, extend arm) in order to perform a set of *target tasks*. The *target tasks* were to pass through a series of obstacles with minimal contact. Similar to the popular American game show "*Hole in the Wall*" (Ludia, 2011). The objective of the applications is to motivate individuals to perform certain physical activities in order to improve their physical health. Moreover, the applications were played in a virtual environment, with participants using full body motion to interact with the virtual environment. Hence, these physically-interactive gamified applications fall within the definition of Active Games (Mears & Hansen, 2009).

After an initial evaluation of the type of movements individuals were going to be required to perform, the Microsoft Kinect sensor was found to be suitable for this study. The Microsoft Kinect was used to capture participants' skeletal data in a non-invasive manner. The sensor is capable of tracking the set of joints shown in Table 3, plus the pelvic joints ($k = 17$). The Microsoft Kinect provides coordinate data of individuals' joints as a distance in meters from a fixed reference point (Behoora & Tucker, 2015). The applications were tested on 71 different participants in a controlled environment in the Design Analysis Technology Advancement (D.A.T.A.) Laboratory at the Pennsylvania State University. The subjects were undergraduate students from the Pennsylvania State University. Their ages ranged from 18 to 23 years old, with a mean of 20 years, and a standard deviation of 1.2 years. From the set of participants, 63 were males, and 8 were females. The objective and the experimental procedure were presented to participants. Subsequently, participants were asked to complete an informed consent document and an online pre-experiment questionnaire that assessed their interest and experience with virtual environments. Finally, participants were randomly assigned to one of the two applications and informed about the objective and how to interact with the gamified application. Due to technical difficulties in the data collection process, only the data from 68 participants were analyzed.

4.1. Initial game features selection

For this study, the information technology system used did not allow participants to play the game more than once ($\alpha = 0$). Furthermore, it was not connected to any network that allowed the system to interact with other similar applications ($\beta = 0$), nor did it allow participants to interact with other players or Artificial Intelligent systems ($\lambda = 0$). These conditions were set for this study in order to minimize the resources required to implement the applications. Following the algorithm shown in Fig. 3 (see section 3.1), it can be shown that due to the constraints of the information technology system used, the set of game features that could be practical to implement were: (i) Avatar, (ii) Boss Fight, (iii) Chance, (iv) Win State, (v) Feedback/Points, (vi) Achievements/Rewards/Badges/Collection, and (vii) Challenges/Content Unlocking/Resource Acquisition/Quest. From the three sets of features that were not mutually exclusive, Points, Achievements, and Content Unlocking were the most suited for the applications used in this study. Additionally, in order to provide empirical evidence to support the results presented by Bharathi et al. (2016), the seven initial features selected were divided into two applications: **A)** with 4 features that were common in successful applications, and **B)** with 3 that were common in unsuccessful applications (see Table 1, section 2.2). The goal of the applications was to motivate individuals to perform certain body motions (e.g., jump, bend, extend arm) in order to perform a set of 12 different *target tasks*. The *target tasks* were to pass through an obstacle with minimal contact.

Table 4
Game features implemented by the applications.

Application A)	Application B)
Points- The score measurement of an individual was shown in the top left corner of his/her visual field. **Avatar-** The individuals were given the option to change the color of the avatar that will represent them in the virtual environment. **Content Unlocking-** Coins were placed throughout the games in different locations. If more than 21 were collected the individual was allowed to change the gaming environment background. **Boss Fight-**At the end of the application there was a very difficult section named "*Boss Fight*".	**Win States-** At the end of the application, the individuals were told if they had won or lost based on a threshold score level. **Chance-**The individuals were given the opportunity to assign a virtual environment background at random. **Achievements-** There were three possible achievements individuals could accomplish shown at the beginning of the application. They were: (i) Lucky Strike: Get through 3 obstacles in a row without touching, (ii) Hops: Jump while going through an obstacle, (iii) Contortionist: Pass every obstacle flawlessly.

The applications were divided into 12 different sections. Each section contained a unique *target task*. Moreover, the two applications shared the same *target tasks* (obstacle avoidance) and only differed on the set of the game features implemented. A total of 37 subjects played the application **A** and 31 played the application **B**. The list of game features and a brief description of how they were implemented in each application is shown in Table 4. The sections of the applications that implemented the feature of Boss Fight were not analyzed since a different set of *target tasks* were implemented which would have affected the comparison analysis between the two applications.

4.2. Game Feature Complexity Analysis

4.2.1. Game feature implementation complexity

The applications were created using Unity (version 5.4). Unity is a cross-platform game engine widely used to develop game applications for computers, consoles, mobile devices, and websites (www.unity3d.com). From the set of initial game features selected, the Implementation Complexity was calculated following the procedure in section 3.2.1. The first step to calculate the Implementation Complexity metric is to identify the information flows. A data flow diagram was created to help visualize the information flows of each of the applications. Fig. 6 shows the diagram of application A and B, respectively. Fig. 6 part A shows the information flows of the game features of Points, Avatar, and Content Unlocking. While part B shows the information flows of the game features of Win State, Chance, and Achievements. Additionally, the obstacles in each application were composed by a set of cubes. Lastly, the game features were integrated via a Game Controller module in both applications.

Table 5 shows the different information flows and the calculated fan-in and fan-out. Moreover, the calculated Implementation Complexity (IC) values for each of the game features implemented in both applications are shown. It can be seen that the game feature of Content Unlocking had the highest Implementation Complexity value. The Content Unlocking feature retrieved information from the global data structures that contain the coins collected and the background environments information. This represents two different Global Flow in ($GF_{in} = 2$) that results in a fan-in of 2 (see Eq. (2)). Lastly, the game feature of Content Unlocking passes information to the Game Controller module. This represents one Local Direct Flow out ($LDF_{out} = 1$) (see Fig. 6, Part A). This resulted in a fan-out of 1 (see

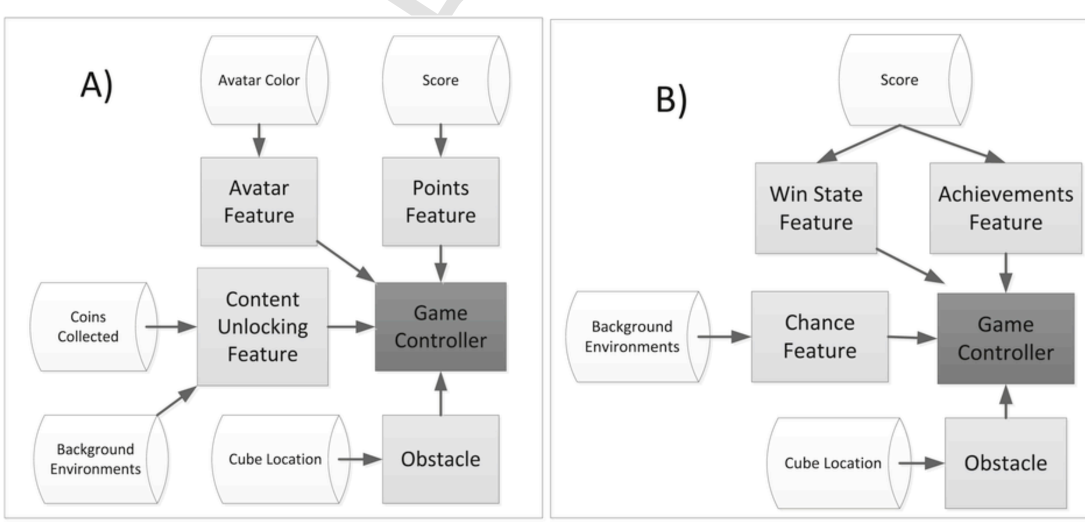

Fig. 6. Data flow diagram of applications A and B.

Table 5
Implementation complexity of applications A and B.

Features	LDF_{in}	GF_{in}	Fan-In	LDF_{out}	GF_{out}	Fan-out	IC
Application A game features IC							
Avatar	0	1	1	1	0	1	**2**
Points	0	1	1	1	0	1	**2**
Content Unlocking	0	2	2	1	0	1	**3**
Application B game features IC							
Win State	0	1	1	1	0	1	**2**
Chance	0	1	1	1	0	1	**2**
Achievements	0	1	1	1	0	1	**2**

Eq. (3)). Following Eq. (1) it can be shown that the Implementation Complexity of the Content Unlocking feature was 3.

4.2.2. Physical Complexity

The Physical Complexity metric was calculated using human skeletal data collected from a pilot test of the applications, following section 3.2.2. From the game features implemented, only the Content Unlocking feature added an *indirect task* to the section of application A. Additionally, the distribution of the coins needed to implement the game feature of Content Unlocking was different for each of the sections of application A. This resulted in a Physical Complexity value for the *indirect tasks* of the Content Unlocking feature that varied depending on the sections of the application A. The features of Points, Avatar, Win State, and Chance did not add any *indirect tasks*. The tasks related to the game feature of Achievements were aligned with the *target tasks* of the application (see Table 4). Hence, it did not add any *indirect tasks*. Therefore, only the Physical Complexity metric of the *target tasks* (obstacle avoidance) was calculated in each section for application B. Since two applications with 12 different sections that contained a unique *target task* were implemented, a total of 24 measurements of Physical Complexity were calculated. Table 6, shows the Physical Complexity values for each of the obstacles that represented a *target task* (PC Obstacle), the Physical Complexity of the *indirect tasks* related to the Content Unlocking feature (PC Content Unlocking), and the total Physical Complexity of each applications sections measured in meters.

Fig. 7, part A, shows a representation of the body joints of an individual in the resting position. While part B shows the body joints of an individual while performing the *target task* (obstacle avoidance) of section 4. This image is from application A, showing a red color Avatar and a "*beach*" background environment. Following Eq. (4), the Physical Complexity for the *target task* shown in Fig. 7 part B can be calculated. For example, the Euclidian distance between the coordinate of the pelvis (shown in yellow) while at rest ($X^{rest} = 0.1m$, $Y^{rest} = 0.5m$, $Z^{rest} = 0m$) and while correctly performing the task ($X^t = 0.1m$, $Y^t = 0.2m$, $Z^t = 0.1m$) was $0.31m$. Adding the Euclidian distances for all the 17 joints (see Eq. (4)) a Physical Complexity value of $3.56m$ is obtained (see Table 6).

Table 6
Physical complexity of applications A and B.

Section No.	PC obstacle	PC content unlocking	Application A	Application B
1	3.68	1.34	5.01	3.68
2	4.01	1.22	5.22	4.01
3	3.66	1.24	4.90	3.66
4	3.56	1.73	5.29	3.56
5	10.13	1.48	11.61	10.13
6	11.70	1.02	12.72	11.70
7	7.00	1.73	8.73	7.00
8	6.95	0.75	7.71	6.95
9	14.37	0.84	15.21	14.37
10	17.58	1.56	19.14	17.58
11	6.23	1.91	8.14	6.23
12	5.90	1.05	6.95	5.90

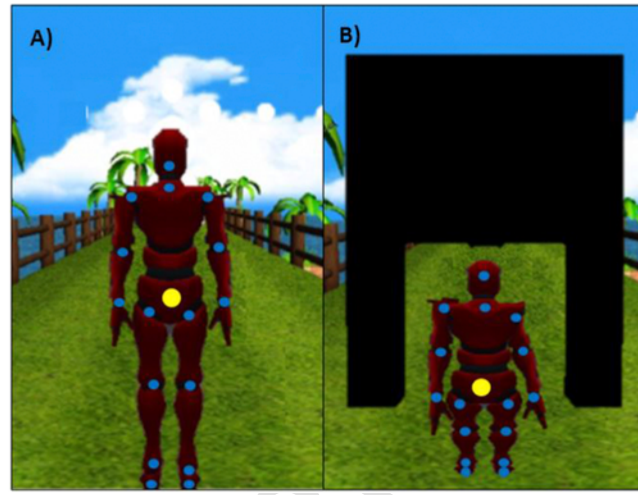

Fig. 7. Representation of the body joints tracked.

4.3. Key Features Discovery

To gain a better understanding of which of the initial game features selected are worth exploring, the Implementation Complexity and Physical Complexity metrics were analyzed. For each feature, the Discovery metric was calculated following Eq. (5) (see section 3.3). Table 8, presents the values for the Discovery metric for each of the game features implemented. The Implementation Complexity (IC) is as shown in Table 5. Furthermore, the Content Unlocking feature was the only one that added *indirect tasks* to the application. On average, the addition of the Content Unlocking feature increased the Physical Complexity (PC) value of the sections of Application A by $1.32m$. The Implementation Complexity and Physical Complexity metrics were normalized following Eq. (5). The Discovery metric suggests that the Content Unlocking feature should not be explored first since it had the highest Implementation Complexity value and was the only feature that added *indirect tasks*.

The distribution of the coins needed to implement the game feature of Content Unlocking varied by section. Hence, the Physical Complexity value for the *indirect tasks* related to this game feature varied by section. This affected the Physical Complexity metric of each section of application A differently (see Table 6). Therefore, the Discovery metrics of Content Unlocking was analyzed conditioned on the sections. The Physical Complexity value for the *indirect task* of each section (PC Content Unlocking, see Table 6), and the normalized Implementation Complexity value for the Content Unlocking feature (Normalized IC, see Table 7) were used to calculate the Discovery metric values conditioned on the sections, shown in Table 8.

Fig. 8 shows a Pareto Chart of the Discovery metric for the Content Unlocking feature conditioned by the sections. It illustrates that in some sections, the Discovery metric is greater than in others (e.g., 11 and 9). Furthermore, the Discovery metric for the Content Unlocking feature for sections 8, 9,6,12 and 2 accounted for less than 20% of the overall accumulative Discovery metric values. This suggests that the Content Unlocking feature should be considered to be implemented as in these sections. Additionally, since the normalized Implementation Complexity value of the Content Unlocking feature was equal to one (see Table 7), the Discovery metric for the Content Un-

Table 7
Game feature discovery metric.

Application	Features	IC	PC	Normalized IC	Normalized PC	Discovery metric
A	Avatar	2	0	0.00	0.00	0.00
	Points	2	0	0.00	0.00	0.00
	Content Unlocking	3	1.32	1.00	1.00	2.00
B	Win State	2	0	0.00	0.00	0.00
	Chance	2	0	0.00	0.00	0.00
	Achievements	2	0	0.00	0.00	0.00

Table 8
Discovery metric for the Content Unlocking feature conditioned by section.

Section No.	Discovery metric
1	0.51
2	0.40
3	0.42
4	0.85
5	0.63
6	0.23
7	0.85
8	0.00
9	0.07
10	0.70
11	1.00
12	0.26

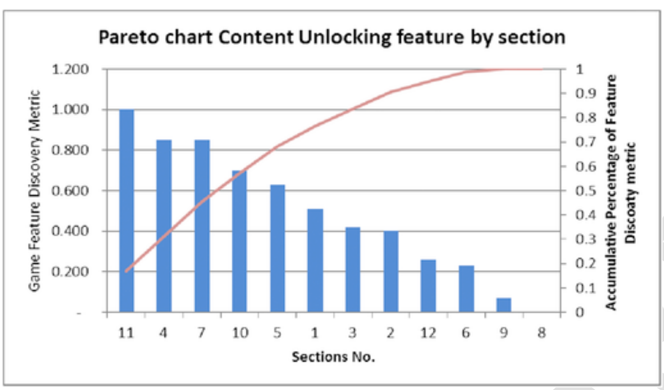

Fig. 8. Pareto chart of the Content Unlocking features discovery metric by section.

locking feature conditioned by the sections directly relates to the Physical Complexity values of the sections.

4.4. Method validation

In this section, quantitative results of individuals' performance are presented in order to provide supporting evidence for the proposed method. The participants' performance was measured by the Intensity of Mistake metric shown in Eq. (6) (see section 3.4). The Wilcoxon-test conducted on the answers provided for the first question of the pre-experiment questionnaire ("*How interested are you in interacting in virtual environments?*", on a 1–5 liker scale) suggests no statistically significant difference in response between participants that played applications A and B at an alpha level of 0.05. Additionally, all the participants were given the same reward (food, beverages, and the opportunity to use an immersed virtual reality system) to be part of the experiment. Based on these results and the relative homogeneous large sample size, the assumption that there was not a significant difference between the initial motivation levels of the participants was made.

4.4.1. Physical Complexity analysis

In this study, the Physical Complexity values for the *indirect tasks* ranged from $0m$ (in the case of application B) to$1.91m$ (section 11, Application A). The Physical Complexity for the *target tasks* ranged from the $3.56m$–$17.58m$ (see Table 6, section 4.2.2). While the Intensity of Mistake values ranged from 0 points (no mistakes) to 6000 points (the individual performed a completely different task than the *target task*). Fig. 9 shows several Box-plots of the distribution of the Intensity of Mistake (IM) conditioned on the Physical Complexity (PC) values of the *indirect tasks* and *target tasks*. In Fig. 9 the Physical Complexity values were rounded to 0 decimal points for visualization proposes. The plots suggest that Intensity of Mistake and Physical Complexity of the tasks are positively correlated. To test the hypothesis that the correlation between Intensity of Mistake and the Physical Complexity of the *indirect tasks* was statistically significantly greater than 0 (see section 3.4, hypothesis 1) a Pearson's product-moment correlation test was conducted in R.v.3.3.0 (Best & Roberts, 1975). The test shows that the correlation is statistically significant at an alpha level of 0.05 ($t = 2.79$, $r = 0.069$). Additionally, the hypothesis that the correlation between Intensity of Mistake and the Physical Complexity of the *target tasks* was significantly greater than 0 (see section 3.4, hypothesis 2) was tested. Similarly, the results

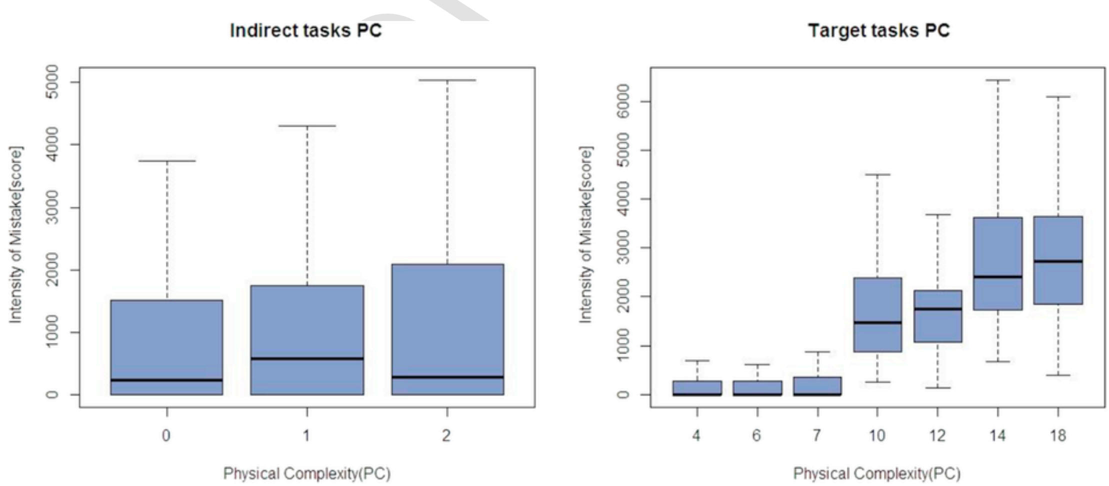

Fig. 9. Box-plots of the distribution of IM given the PC of the indirect tasks and target tasks.

show that the correlation was statistically significant at an alpha level of 0.05 ($t = 33.9$, $r = 0.643$).

These results suggest a significant correlation between individual's performance and the physical effort required to perform the different tasks of the applications. The figures show that the individuals performed better when the Physical Complexity values of the *indirect tasks* and *target tasks* were small. These results support the Discovery metric proposed in this study (see section 3.3) and present quantitative evidence that supports the principle of simplicity shown in several of the models and theories applied to gamified applications (see section 2.1). Additionally, these results are in line with Fogg's Behavioral Model "*behavior activation threshold*" concept. This threshold suggests that if an individual does not possess the ability to perform a task, no matter the stimuli presented to them, they won't be able to perform it (Fogg, 2009). This threshold can be observed in Fig. 9, where the performance of participants does not change significantly until the Physical Complexity value of the *target tasks* exceeds 7*m*.

4.4.2. Application analysis

To provide empirical evidence to support the results presented by Bharathi et al. (2016), the initial features selected were divided into two different applications (see section 4.1). Application A implemented the features that were common across successful applications, while application B implemented the features that were common across unsuccessful applications. To test the hypothesis that there was a significant difference in performance between participants that interact with application A vs application B, a two-sided *t*-test was performed (see section 3.4, hypothesis 3). The *t*-test suggests that the difference in the average Intensity of Mistake between participants that interacted with application A vs application B was statistically significant at an alpha level of 0.05 ($t = 2.82$). The results suggest that individuals that interacted with Application A on average, performed worse than the ones that interacted with Application B ($\mu_{IM|A}$=1141 points, $\mu_{IM|B}$ = 934 points). This can be attributed to the fact that the physical effort required to perform the sections of application A were greater than the ones on application B. This can be seen in Table 6 were the Physical Complexity values for the sections of application A are greater than application B, due to the addition of the *indirect tasks* product of the Content Unlocking feature.

Since the Physical Complexity values of the *indirect tasks* related to the Content Unlocking feature varied by section, the participants' performance was analyzed conditioned on the sections. Fig. 10 shows the main effects of the application type and sections on the average Intensity of Mistake. From these effect plots, it is clear that on average, participants performed better in sections: 1, 2, 3, 4, 7, 8, 11, and 12 than in sections: 5, 6, 9, and 10. This can be attributed to the physical effort required to perform the *target task* in those sections. The Physical Complexity values for the *target task* of sections 5, 6, 9, and 10 ranged from 10.13*m* to 17.58*m*. While for the other sections it ranged from 3.56*m* to 7*m* (see Table 6, section 4.2.2). Additionally, in Fig. 10, some interaction effects between the application type and the section can be observed. In some sections, the average Intensity of Mistake decreases when moving from application A to B (e.g., section 5, 9, 11, and 12), while for others sections it seems to increase (e.g., section 2, 6, and 8).

To test the hypothesis that there is a significant difference in performance between participants that interact with application A vs application B when conditioned by the sections, a series of two-sided t-tests were conducted (H$_o$: $IM_A|s = IM_B|s$; H$_a$: $IM_A|s \neq IM_B|s$, for s ε Sections {1–12}). Table 9 shows the t-statistics of these tests alongside the average Intensity of Mistake for each application conditioned by section. These results suggest that participants that interacted with Application A, on average performed statistically significantly worse compared to participants that interacted with application B on sections 1, 3, 5, 11, and 12. Even though the average Intensity of Mistake for application A for section 2 and 8 is lower than application B, it was not statistically significant at an alpha level of 0.05. Nonetheless, the results suggest that participants who interacted with Application A on average performed statistically significantly better on section 6 compared to participants that interacted with application B. These results can be attributed to the interaction between the physical effort required to perform the *target task* and the *indirect task* on

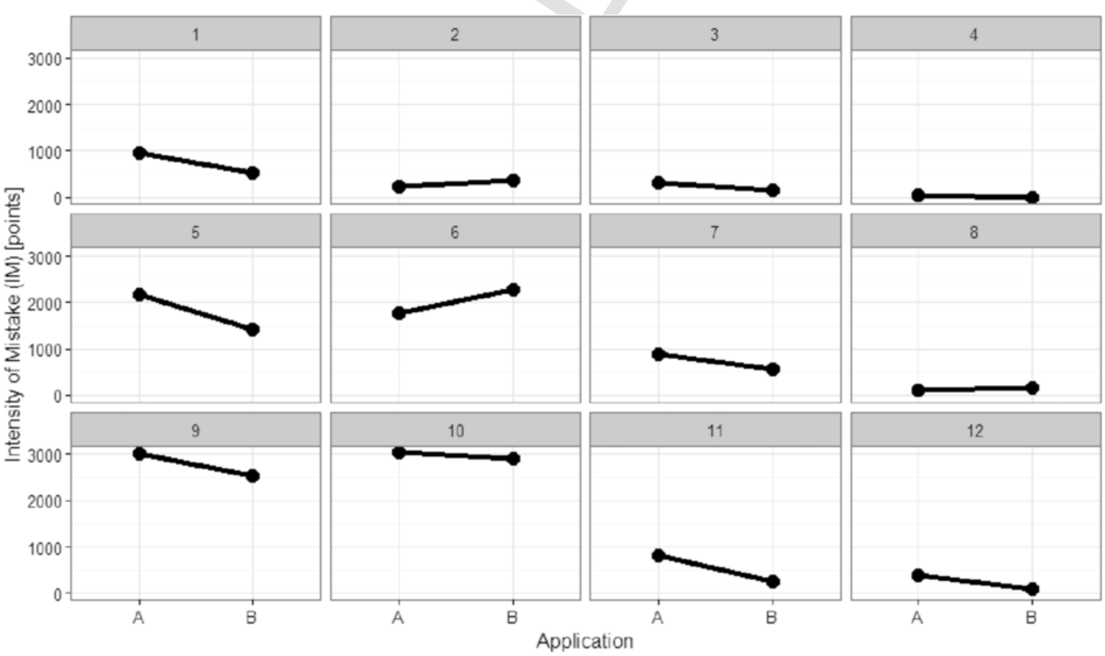

Fig. 10. Interaction plot of the average Intensity of Mistake given Application (A vs B) and Sections (1–12).

Table 9
T-test conditioned by section.

Section	t-statistic	Average IM_A	Average IM_B
1	2.371*	959.51	518.92
2	−1.042	232.63	364.69
3	2.153*	297.99	139.73
4	1.828	34.48	0.00
5	3.439*	2179.32	1412.61
6	−2.137*	1763.28	2268.77
7	1.823	887.07	562.70
8	−0.369	111.47	164.12
9	1.922	3007.91	2523.98
10	0.457	3034.32	2904.98
11	2.861*	802.88	253.20
12	2.89*	381.74	98.72

* Statistical significance at an alpha level of 0.05.

these sections. Table 6 (see section 4.2.2) shows that the *indirect tasks* related to the Content Unlocking feature have a higher Physical Complexity value for sections 1, 3, 5, 11, and 12 (1.34*m*, 1.24*m*, 1.48*m*, 1.91*m*, 1.05*m*, respectively) than for section 6 (1.02 m). Moreover, Fig. 8 shows that sections 2, 6, 8, 9, and 12 accounted for less than 20% of the overall accumulative Physical Complexity values of application A. However, the Physical Complexity value of the *target tasks* of sections 5 and 6 are higher than for sections 1, 3, 11, and 12 (see Table 6, section 4.2.2). These results suggest that the effects of the additional physical effort required to perform the *indirect tasks* related to the Content Unlocking feature may be mediated by the physical effort required to perform the *target tasks* of the applications. These results are in line with Fogg's Behavioral Model. The model states that there exist a relationship between individuals' motivation and ability to perform a task (Fogg, 2009). If the ability of an individual to perform a task or the simplicity of the task is reduced, while his/her motivation level does not change, the performance of the individual will deteriorate. Furthermore, the effects of the additional *indirect tasks* related to the Content Unlocking feature did not have a significant negative effect at an alpha level of 0.05 on the performance of participants on 7 out of the 12 sections. These results suggest that the game features of Points and Avatar mitigated the negative effect of the additional *indirect tasks* on the performance of individuals. These results are in line with previous studies that have shown positive results by implementing the game features of Points and Avatar (see section 2.2). Additionally, it provides empirical evidence to support the results presented by Bharathi et al. (2016).

4.4.3. Summary of results

In summary, the results show that:

1. Individuals' performance is strongly correlated to the physical effort required to perform a *target task*.
2. There exist threshold levels of physical effort for which individuals' performance on a *target task* will decrease significantly.
3. Individuals' performance on a *target task* can be negatively affected by an *indirect task* resulting from the implementation of a game feature.
4. The effect of a game feature on individuals' performance will depend on the physical effort required to perform any *indirect task* related to its implementation as well as the physical effort required to perform the *target task* of the application.
5. The game features of Points and Avatar have a positive effect on the performance of individual in comparison with the game features of Win State, Chance, and Achievements.
6. The game features of Points and Avatar mitigated the negative effects of the additional *indirect tasks*, product of the Content Unlocking feature, on the individuals' performance.

By applying the proposed method, a better understanding of the effects of game features and the physical effort required to perform them was gained. The results suggest a strong correlation between participants' performance and the physical effort required to perform the *target tasks* and *indirect tasks* of the physically-interactive gamified applications analyzed. These results provide quantitative evidence that supports the principle of simplicity which is present in several of the models and theories of gamified application (see section 2.2). Moreover, the results provide quantitative data in support of the proposed method. This method can help designers to quantitatively evaluate the effect that game features have on individuals' performance. Furthermore, the results of the case study suggest that effects of game features on participants' performance will depend on the physical effort required to perform the *indirect tasks* related to the game feature, and the *target tasks* of the application. This knowledge can inform designers on how to manage their creative resources more efficiently.

5. Conclusion and future work

This study presented a method that quantitatively explores the complexity of both implementing and performing game features in physically-interactive gamified applications. This method will help designers to gain a fundamental understanding of the effects of game features on individuals' performance. The method provides a systematic approach for analyzing game features that are worth exploring and implementing in physically-interactive gamified applications. An initial game feature selection algorithm based on previous studies was proposed. This algorithm will allow designers to gain a better understanding of which game features are practical to implement based on the constraints of the information technology systems used for their applications. Furthermore, the complexity of implementing game features and the physical effort required to perform *indirect tasks* related to game features are used to assess which features are worth exploring. The method was tested on a physically-interactive gamified application in a virtual environment. The objective of the applications is to motivate individuals to perform certain physical activities in order to improve their physical health. The *target tasks* of the gamified applications were to pass through a series of obstacles with minimal contact. In this study, the Microsoft Kinect was used to acquire human skeletal data. This data was used to measure the physical effort required to perform the *target tasks* of the application, and the *indirect tasks* of the applications. A Physical Complexity metric was presented that allows designers to capture the physical effort required to perform a task in a physically-interactive gamified application. Quantitative results of the effect of well known game features on the performance of individuals for applications in a virtual environment were presented. This expanded the current body of knowledge gamification that was in need of more empirical results on the usefulness of gamified applications in improving user engagement and performance. The results suggest that the physical effort required to perform *indirect tasks* and *target tasks* have a significant effect on participant's performance. These results support the method proposed that focused on minimizing the physical effort needed to perform any *indirect task* related to the implementation of a game feature. Moreover, the results suggest that the effects of game features on individuals' performance will be mediated by the physical effort required to perform the *indirect tasks* and the *target tasks* of the application. It provided quantitative results supporting the principle of simplicity, the relationship between individuals' performance and the physical effort required to perform a task, and the "*behavior activation threshold*" present in Fogg's Behavioral Model. Additionally, the results suggested that the implementation of the game features of Points and Avatar had a positive effect on individuals' performance. These results provide empirical evidence that supports the finding of the previous research, such as Bharathi et al. (2016) that found this set of features in successful driven applications. Similarly, the findings of Dicheva et al. (2015), Pedreira et al. (2015), and Hamari et al. (2014) found that in the field gamification research the features of Points and Avatar have been shown to have positive results on improving individual engagement. The Physical Complexity metric proposed in this method was significantly correlated with the participants' performance, showing that with this metric individual's performance in physically-interactive gamified applications can be predicted. Nonetheless, the use of other variables alongside the Physical Complexity metric might be needed to improve performance predictions. One of the limitations of this study was that information about the physical fitness of the participants was not collected prior to the experiments, which could have affected their performance. Furthermore, participants' performance was assumed to be related to their motivational levels. However, the physical fitness of the participants might be an important mediator variable in this relationship. Future works should explore other variables that might play an important role in predicting individuals' performance (e.g., physical fitness conditions and personality traits). Similarly, other metrics that allow the individuals' motivational level to be measured should be considered. Nonetheless, this method can be applied to existing physically-interactive gamified applications to understand their weaknesses and strengths of their design and to provide arguments for their results. This method will help to inform designers on how to manage their creative resources more efficiently and guide them in the selection of game features to implement.

Acknowledgements

This research was funded in part by NSF DUE #1449650 and NSF NRI #1527148. Any opinions, findings, or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors. The authors would also like to thanks the support of Abhinav Singh, Matt Young, Sam Barnes and the participants for their help in this study.

Appendix A.

Mathematical Proof

Under the assumptions:

1) Task λ that requires the individual to perform a full body motion (e.g., jumping, bending, walking) can be accurately represented by the set of joints presented in Table 3 (this assumption is based on previous studies, see section 3.2.2)
2) The sensor 1 and 2 use the same distance unit.

Under the conditions:

1) Tracking sensor 1 (S1) is capable of only tracking the set of joints shown in Table 3, (set **K)**
2) Tracking sensor 2 (S2) is capable of tracking the joints shown in Table 3 plus 2 more joints, (set **D,** where **D** = **K**∪***E***, and **E**:{joint16,joint17})

It can be proven that the Physical Complexity metric of game feature λ using tracking sensor 2 ($PC_{\lambda|S2}$) can be greater or equal than the Physical Complexity metric of game feature λ using tracking sensor 1 $PC_{\lambda|S1}$. Therefore:

$$PC_{\lambda|S1} \leq PC_{\lambda|S2}$$

Proof:
From Eq. (4):

$$PC_{\lambda|S1} = \sum_{j=1}^{\boldsymbol{K}} \sqrt[2]{\left(X_j^{rest} - X_j^t\right)^2 + \left(Y_j^{rest} - Y_j^t\right)^2 + \left(Z_j^{rest} - Z_j^t\right)^2} \quad \text{for } t = \lambda$$

$$PC_{\lambda|S2} = \sum_{j=1}^{\boldsymbol{D}} \sqrt[2]{\left(X_j^{rest} - X_j^t\right)^2 + \left(Y_j^{rest} - Y_j^t\right)^2 + \left(Z_j^{rest} - Z_j^t\right)}$$

For simplicity let's rewrite the equations as:

$$XYZ_{js}^2 = \left(X_j^{rest} - X_j^t\right)^2 + \left(Y_j^{rest} - Y_j^t\right)^2 + \left(Z_j^{rest} - Z_j^t\right)^2 \ \ for \ \ s \in Set\ of\ sensors\ \{S1,\ S2\}$$

Under assumption 2:

$$XYZ_{js1}^2 = XYZ_{js2}^2 = XYZ_j^2 \ \ for\ j \ \in \boldsymbol{K}$$

Since:

$$\boldsymbol{D} = \{\ \boldsymbol{K} \cup \boldsymbol{E}\}$$

$$PC_{\lambda|S1} = \sum_{j=1}^{\boldsymbol{K}} \sqrt[2]{XYZ_j^2}$$

$$PC_{\lambda|S2} = \sum_{j=1}^{\boldsymbol{K}} \sqrt[2]{XYZ_j^2} + \sum_{j=1}^{\boldsymbol{E}} \sqrt[2]{XYZ_{js2}^2}$$

Under assumption 1:

$$\sum_{j=1}^{\boldsymbol{E}} \sqrt[2]{XYZ_{js2}^2} = 0$$

Then:

$$PC_{\lambda|S1} = PC_{\lambda|S2}$$

If assumption 1 is violated:

$$\sum_{j=1}^{E} \sqrt[2]{XYZ_{js2}^{2}} \neq 0$$

Then:

$$PC_{\lambda|S1} \leq PC_{\lambda|S2}$$

This shows that under assumption 1 and 2 if a sensor that is capable of tracking more joints than the one listed in Table 3, the accuracy of the Physical Complexity can only be greater than in the case that a sensor that is capable of tracking only the joints shown in Table 3 is used.

References

Aggarwal, K.K., Singh, Y., 2005. Software engineering programs documentation, operating procedures (Revised Se). New Age International Publishers.

Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J., 2014. Engaging with massive online courses. In: WWW'14 proceedings of the 23rd international conference on world wide web. pp. 687–698. http://dx.doi.org/10.1145/2566486.2568042.

Barata, G., Gama, S., Jorge, J., Gonçalves, D., 2016. Studying student differentiation in gamified education: A long-term study. Computers in Human Behavior

Behoora, I., Tucker, C.S., 2015. Machine learning classification of design team members' body language patterns for real time emotional state detection. Design Studies 39, 100–127. http://dx.doi.org/10.1016/j.destud.2015.04.003.

Best, D.J., Roberts, D.E., 1975. Algorithm AS 89: The Upper Tail Probabilities of Spearman's rho. Applied Statistics 24, 377–379.

Bharathi, G.B.A.K., Singh, A., Tucker, C.S., Nembhard, H.B., 2016. Knowledge discovery of game design features by mining user-generated feedback. Computers in Human Behavior 60, 361–371. http://dx.doi.org/10.1016/j.chb.2016.02.076.

Biddiss, E., Irwin, J., 2010. Active video games to promote physical activity in children and youth. Archives of Pediatric and Adolescent Medicine 164 (7), 664–672. http://dx.doi.org/10.1001/archpediatrics.2010.104.

Brauner, P., Calero Valdez, A., Schroeder, U., Ziefle, M., 2013. Increase physical fitness and create health awareness through exergames and gamification. The role of individual factors, motivation and acceptance. In: Proceedings of the SouthCHI 2013. LNCS, pp. 349–362. http://dx.doi.org/10.1007/978-3-642-39062-3_22. 7946, (August).

Caton, H., Greenhill, D., 2013. The effects of gamification on student attendance and team performance in a third-year undergraduate game production module. In: 7th European conference on games based learning, ECGBL. pp. 88–96.

Chan, C.K., Loh, W.P., Rahim, I.A., 2016. Human motion classification using 2D stick-model matching regression coefficients. Applied Mathematics and Computation 283, 70–89. http://dx.doi.org/10.1016/j.amc.2016.02.032.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper & Row, New York.

Denny, P., 2013. The effect of virtual achievements on student engagement. In: ACM CHI'13. pp. 763–772.

Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From game design elements to Gamefulness: Defining " gamification.". In: ACM MindTreck'11.

Dicheva, D., Dichev, C., Agre, G., Angelova, G., 2015. Gamification in Education : A systematic mapping study. Educational Technology & Society 18 (3), 75–88.

Fogg, B.J., 2009. A behavior model for persuasive design. In: Proceedings of the 4th international conference on persuasive technology. ACM, New York, NY, USA, pp. 40:1–40:7. http://dx.doi.org/10.1145/1541948.1541999.

Gao, Z., Chen, S., 2014. Are field-based exergames useful in preventing childhood obesity? A systematic review. Obesity Reviews 15 (8), 676–691. http://dx.doi.org/10.1111/obr.12164.

Hamari, J., Koivisto, J., Sarsa, H., 2014. Does gamification work? - A literature review of empirical studies on gamification. In: Proceedings of the annual Hawaii international conference on system sciences. pp. 3025–3034. http://dx.doi.org/10.1109/HICSS.2014.377.

Han, F., Reily, B., Hoff, W., Zhang, H., 2016. Space-time representation of people based on 3D skeletal data: A review. arXiv Preprint arXiv:1601.01006.

Henry, S., Selig, C., 1990. Predicting source-code complexity at the design stage. IEEE Software 7 (2), 36–44.

Jabangwe, R., Börstler, J., Šmite, D., Wohlin, C., 2015. Empirical evidence on the link between object-oriented measures and external quality attributes: A systematic literature review. Empirical Software Engineering 20 (3), 640–693.

Jibb, L., Stinson, J., Nathan, P., Maloney, A., Dupuis, L., Gerstle, T., … Yanofsky, R., 2012. Pain squad: Usability testing of a multidimensional electronic pain diary for adolescents with cancer. The Journal of Pain 13 (4), S23.

Li, W., Grossman, T., Fitzmaurice, G., 2012. GamiCAD: A gamified tutorial system for first time AutoCAD users. In: ACM UIST'12. pp. 103–112.

Linehan, C., Kirman, B., Lawson, S., Chan, G.G., 2011. Practical, appropriate, empirically-validated guidelines for designing educational games. In: CHI '11 proceedings of the SIGCHI conference on human factors in computing systems, (July 2015), 1979–1988. http://dx.doi.org/10.1145/1978942.1979229.

Lucassen, G., Jansen, S., 2014. Gamification in consumer marketing - future or fallacy?. Procedia Social and Behavioral Sciences International Conference on Strategic Innovative Marketing 148 (2011), 194–202. http://dx.doi.org/10.1016/j.sbspro.2014.07.034.

Ludia, 2011. Ludia & Fremantle Media enterprises ink deal to bring hole in the WallTM games exclusively for KinectTM on Xbox 360(r). Electronics Business Journal 64.

Malone, T.W., 1982. Heuristics for designing enjoyable user interfaces: Lessons from computer games. In: Proceedings of the 1982 conference on human factors in computing systems. pp. 63–68. http://dx.doi.org/10.1145/800049.801756.

McCallum, S., 2012. Gamification and serious games for personalized health. pHealth 85–96. http://dx.doi.org/10.3233/978-1-61499-069-7-85.

Mears, D., Hansen, L., 2009. Active Gaming: Definitions, options and implementation. Article #5 in a 6-Part Series Strategies: A Journal for Physical and Sport Educators 23 (2), 26–29.

Mens, T., 2016. Research trends in structural software complexity. arXiv Preprint, arXiv:1608.01533.

Oinas-Kukkonen, H., Harjumaa, M., 2008. A systematic framework for designing and evaluating systems. In: Persuasive technology, third international conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008. pp. 164–176. http://dx.doi.org/10.1007/978-3-540-68504-3. Proceedings, 5033.

O'Donovan, S., Gain, J., Marais, P., 2013. A case study in the gamification of a university-level games development course. SAICSIT 242–251. http://dx.doi.org/10.1145/2513456.2513469. East London, South Africa.

Pedreira, O., Garcia, F., Brisaboa, N., Piattini, M., 2015. Gamification in software engineering - a systematic mapping. Information and Software Technology 57 (1), 157–168. http://dx.doi.org/10.1016/j.infsof.2014.08.007.

Phukan, A., Kalava, M., Prabhu, V., 2005. Complexity metrics for manufacturing control architectures based on software and information flow. Computers & Industrial Engineering 49 (1), 1–20. http://dx.doi.org/10.1016/j.cie.2005.01.005.

Read, J.L., Shorter, S.M., 2011. Interactive games to promote behavior change in prevention and treatment. American Medical Association JAMA 305 (16), 1704–1705. http://dx.doi.org/10.1001/jama.2011.408.

Rose, K.J., Koenig, M., Wiesbauer, F., 2013. Evaluating success for behavioral change in diabetes via mHealth and gamification: MySugr's keys to retention and patient engagement. Diabetes Technology & Therapeutics 15. A114–A114.

Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55 (`), 68–78.

Sell, K., Lillie, T., Taylor, J., 2008. Energy expenditure during physically interactive video game playing in male college students with different playing experience. Journal of American College Health: J of ACH 56 (5), 505–511. http://dx.doi.org/10.3200/JACH.56.5.505-512.

Shen, J., Yang, W., Liao, Q., 2011. Multiview human pose estimation with unconstrained motions. Pattern Recognition Letters 32 (15), 2025–2035. http://dx.doi.org/10.1016/j.patrec.2011.09.019.

Stinson, J.N., Jibb, L.A., Nguyen, C., Nathan, P.C., Maloney, A.M., Dupuis, L.L., … Donna, L., 2013. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. http://dx.doi.org/10.2196/jmir.2350. Corresponding Author, 15.

Werbach, K., Hunter, D., 2012. For the Win: How game thinking can revolutionize your business. Wharton Digital Press.

Xiao, Z., Nait-Charif, H., Zhang, J.J., 2008. Automatic estimation of skeletal motion from optical motion capture data. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5277 LNCShttp://dx.doi.org/10.1007/978-3-540-89220-5-15.

Yang, J.C., Chen, C.H., Chang, M., 2010. Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning. Computers and Education 55 (3), 1346–1356. http://dx.doi.org/10.1016/j.compedu.2010.06.005.