Computers in Human Behavior 97 (2019) 151–166

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

"Achievement unlocked!" - The impact of digital achievements as a gamification element on motivation and performance

Christopher Groening*, Carmen Binnewies

Work Psychology Unit, Institute of Psychology, University of Muenster, Germany

ARTICLE INFO

Keywords:
Gamification
Achievements
Goal-setting
Motivation
Performance
Experiment

ABSTRACT

Gamification experiences further growth in our society with broad practical implications of game design elements in applications, activities, and services. In our study, we focus on one single element in the form of digital achievements, one cornerstone of gamification, to gain unconfounded insights into the effects and working mechanisms of digital achievements. In a controlled experimental environment we investigate their impact on motivation and performance. Three research questions were investigated, first, if achievements are effective in enhancing motivation and performance, second, how achievements need to be designed to be effective, and third, if achievements' underlying working mechanisms resemble classical goal-setting. In three experiments, a total of 245 participants worked on different cognitive tasks. We experimentally manipulated several aspects of achievements, such as quantity and difficulty. We also compared achievements directly with classical goal-setting instructions. Results showed that achievements enhanced performance. Achievements also improved motivation, however only regarding persistence, not self-reported interest and enjoyment. The effectiveness of achievements was highly dependent on the design. Results suggest utilizing achievements with high difficulty and in low quantity. Results also support our assumption that achievements provide a direct goal-setting function. We conclude that achievements do benefit motivation and performance if designed properly.

1. Introduction

Do you know a gamer? Maybe you are a gamer yourself. If not, the chances are high that there will be an upcoming change soon. The growing spread of gaming and video games has reached a point, at which it can be seen as a significant part of society. Sixty-five percent of American households are home to someone who plays video games regularly (Entertainment Software Association, 2017). With the average player being 35 years old, and one out of four gamers being over the age of 50 (Entertainment Software Association, 2017), even CEOs, CFOs, and other senior executives take daily game breaks at work (Reinecke, 2009). The motivational potential of video games has been addressed and found in several psychological studies and literature (e.g., Dickey, 2006; Garris, Ahlers, & Driskell, 2002; Granic, Lobel, & Engels, 2014; Peng, Lin, Pfeiffer, & Winn, 2012; Przybylski, Rigby, & Ryan, 2010; Yee, 2006). The potential of motivational benefits of games and gaming has been seized by the industry outside the branch of classical video games. A new concept was—and still is—pitched as a next-generation tool for originally marketing, consumer retention, or user engagement (Hamari, Koivisto, & Sarsa, 2014). This concept was coined as *gamification* and is known as the "use of game design elements in non-game contexts" (Deterding, Khaled, Nacke, & Dixon, 2011, p.1). The central idea of gamification is to transfer the motivational potential of video games to non-game environments. However, a simple transfer of game elements is usually seen as not sufficient for labeling this process "gamification" (Werbach, 2014). Gamification shall create an activity which evokes game-like experiences. Sailer, Hense, Mayr, and Mandl (2017) proposed a more nuanced definition of gamification: "the process of making activities in non-game contexts more game-like by using game design elements" (p. 372). Our paper presents three experiments, contributing to gamification research of achievements, motivation, and performance.

Within recent years, gamification has become a trending topic (Hamari et al., 2014). Analysts predict a future constant growth with an estimated increase from $1.65 billion in 2015 to $11.1 billion by 2020 at a compound annual growth rate (CAGR) of 46.6% (Markets and Markets, 2016). The use of gamification is very manifold and covers a wide spectrum of anticipated benefits from the initiators' point of view. For example, gamification is used to support user engagement and user activity (e.g., Hamari et al., 2014), environmental behavior, motivating for physical workout, fostering safe drive behavior (e.g., Sailer, Hense, Mandl, & Klevers, 2013), work performance (e.g., Fernandes et al.,

* Corresponding author.
E-mail address: christopher.groening@uni-muenster.de (C. Groening).

https://doi.org/10.1016/j.chb.2019.02.026
Received 5 November 2018; Received in revised form 25 February 2019; Accepted 26 February 2019
Available online 01 March 2019

2012; Mollick & Rothbard, 2013) or education (e.g., Landers, 2014; Muntean, 2011). As the definitions above suggest, the basic concept of gamification comprises transferring game design elements of a (video) game to a non-game context, where it is hoped to have similar beneficial behavioral and motivational outcomes (amongst others) as games. We refer to game design elements as the description of elements that are characteristic to games (for a more detailed discussion about the definition of game elements see Deterding, Dixon, Khaled, & Nacke, 2011). Game design elements deployed in gamification are numerous (for a comprehensive overview see Seaborn & Fels, 2015; Werbach & Hunter, 2012), whereas a majority of gamification systems are built upon a foundation of three game elements: (a) points, (b) badges, (c) leaderboards (Hamari et al., 2014; Seaborn & Fels, 2015; Werbach & Hunter, 2012).

Points are a basic element of games (Zicherman & Cunningham, 2011), as they serve as a feedback mechanism (Werbach & Hunter, 2012) and as a reward (Sailer et al., 2013). Badges, usually seen as a visual representation of an achievement, are secondary reward systems, where users complete sub-goals to earn achievement rewards (Montola, Nummenmaa, Lucero, Boberg, & Korhonen, 2009). A leaderboard represents an ordered display of scores, measured against a particular success criterion, together with the names of participants.

Empirical findings on gamification have been mixed (Hamari et al., 2014; Koivisto & Hamari, 2019). For example, positive effects have been found for motivation (e.g., Denny, 2013; Su & Cheng, 2015), performance (e.g., Mollick & Rothbard, 2013; Su & Cheng, 2015; Witt, Scheiner, & Robra-Bissantz, 2011), persistence (e.g., Landers & Landers, 2015), user activity (e.g., Anderson, Huttenlocher, Kleinberg, & Leskovec, 2013) or contribution (e.g., Farzan et al., 2008). Other studies revealed no effects, effects constrained to certain conditions, or adverse effects (e.g., Attali & Arieli-Attali, 2015; de Marcos, Domínguez, de Navarrete, & Pagés, 2014; Guin, Baker, Mechling, & Ruyle, 2012; Hamari, 2013; Hanus & Fox, 2015; Koivisto & Hamari, 2014). Hamari et al. (2014) concluded in their literature review that gamification does work, but some caveats exist. They point out that the context of service might be an essential antecedent for engaging gamification, such as the user involvement, the social environment, or the nature of the system. Regarding the latter, we attribute the mixed findings of gamification literature especially to the diverse design of the applied game design elements and gamification systems. They vary in numerous dimensions, for example, quantity of elements (e.g., amount of achievements), variety of elements (e.g., points, leaderboards, graphs, narrative, badges), visual presentation (e.g., animations, appeal), difficulty (e.g., easy achievements), usage time (e.g., single task, long-term usage).

In our study, we focus on one specific element, namely digital achievements. In three experiments, we investigate their effects on motivation and performance. This allows for an unconfounded investigation of game design elements by taking only one specific element into focus, enabling specific propositions of the effectiveness of that specific game design element. We also deployed tasks and environments which were as free of context as possible. This approach contributes to more generally applicable propositions about the effects of gamification.

2. Goal-setting theory

Authors in gamification research have proposed that achievements exert a goal-setting function (Antin & Churchill, 2011; Sailer et al., 2013; Werbach & Hunter, 2012). Achievements—as a secondary reward-system—encourage users to show a proposed behavior to receive a specific achievement as a reward. We also propose that achievements evoke similar mechanisms as classical goal-setting, which usually is instructed in a clear and direct way (e.g., “Your goal in this task is to solve 15 mathematical problems.“) along the *goal-setting theory* (Locke & Latham, 1990).

Based on early findings of relationships between conscious goals and performance (Locke, 1968), Locke and Latham (1990) provide a well-developed goal-setting theory of work motivation. A goal can be understood as the aim of an action, which the individual tries to accomplish, for example, the strive to attain a certain proficiency level within a specified time limit (Locke & Latham, 2002). A goal can be self-set or externally provided by supervisors or peers, for example. Under the right conditions, setting a goal may be a powerful technique to enhance motivation and performance (Lunenburg, 2011). In 2006, Locke and Latham summarized that over a 25-year period 400 laboratory and field studies “showed that specific, high (hard) goals lead to a higher level of performance than do easy goals or vague, abstract goals such as the exhortation to ‘do one's best’.” (p. 265). According to Locke, Shaw, Saari, and Latham (1981), goals influence motivation directly. Performance is affected through four mechanisms (Locke & Latham, 2002). First, goals serve a directive function directing attention and effort toward goal-relevant activities, both cognitively and behaviorally (Locke & Bryan, 1969; Rothkopf & Billington, 1979). Second, high goals provide an energizing function, leading to greater effort than low goals (Bryan & Locke, 1967). Third, goals affect persistence where hard goals prolong the effort invested (LaPorte & Nath, 1976). Fourth, goals indirectly affect action by leading to the discovery and use of task-relevant knowledge and strategies (Wood, 1990). Well-known important moderators of the effectiveness of goal-setting are ability, goal commitment, task complexity, and situational constraints, which can enhance or hinder the goal-setting effect on motivation and performance (Latham, 2015). Goal-setting has been shown as a very robust effect in psychological literature with meta-analyses providing effect sizes (standardized mean differences *d*) ranging from 0.42 to 0.80 (Locke & Latham, 1990).

3. Research objectives

The focus of the experiments in this work lies on digital achievements. They are the cornerstone of gamification as the vast majority of gamification systems incorporate achievements as a game design element (Hamari et al., 2014; Seaborn & Fels, 2015; Werbach & Hunter, 2012). Therefore, achievements should be considered high priority when investigating game design elements. Several studies investigated achievements (e.g., Barata, Gama, Jorge, & Goncalves, 2013; Domínguez et al., 2013; Donato & Michael, 2013; Hamari, 2013; Koivisto & Hamari, 2014; Liu, Alexandrova, & Nakajima, 2011; McDaniel, Lindgren, & Friskics, 2012; Sailer et al., 2017; for an overview see also Seaborn & Fels, 2015). However, controlled experimental studies, examining the effects of one game design element at a time, are still scarce (Koivisto & Hamari, 2019), especially in which achievements were deployed (Hakulinen, Auvinen, & Korhonen, 2015, 2013; Hamari, 2013). We aim to combine the benefits of an unconfounded single-element approach with a preferably context-free laboratory experimental environment to gain new insights about the effectiveness of achievements as an essential game design element in gamification.

In this work, we focus on performance and motivational benefits as a major outcome of gamification. Many authors of empirical studies and theoretical papers assume that one essential outcome of gamification is the motivational benefit (e.g. Aparicio, Vela, Sanchez, & Montes, 2012; Hanus & Fox, 2015; Hense et al., 2014; Nicholson, 2012). Furthermore, several performance indicators as outcomes of gamification have been one of the main investigated factors in gamification literature (e.g. Arai, Sakamoto, Washizaki, & Fukazawa, 2014; Attali & Arieli-Attali, 2015; Landers & Landers, 2015; Mekler, Brühlmann, Opwis, & Tuch, 2013; Mollick & Rothbard, 2013). As literature is mixed about the question if achievements foster motivation and performance, our first main research objective is to gain further insights into the question if achievements are effective in enhancing motivation and performance.

As a second research objective, we intend to address design aspects of achievements. As we stated earlier, we believe that controversial

results in gamification research are also due to the highly diverse design of the applied game design elements. Most of gamification literature has neglected individual aspects of gamification design. For example, imagine two gamification systems, both include achievements. Due to their distinctive design (quantity, difficulty, visual appeal, contextual framing, etc.), user experience may be highly different. However, usually, studies made merely a dichotomous distinction of achievement—present versus absent—neglecting design-related distinctions. We address this issue by comparing different designs of achievement systems experimentally to identify potential key dimensions which may be important regarding achievement effectiveness. We focus on the aspects of difficulty and quantity of achievements, as they are inherent aspects of every achievement system.

We propose that goal-setting is a major underlying psychological working mechanism of the motivational benefits of achievements. In their literature review Seaborn and Fels (2015) emphasized a disconnection between theoretical and applied works on gamification. There exists theoretical work explaining the potential inner workings of gamification. However, Seaborn and Fels (2015) state that it "has not been empirically validated with respect to applied gamification work. Likewise, applied work may reference theory but does not explore its validity empirically." (p. 29). We investigate achievements drawing upon the goal-setting theory (Locke & Latham, 1994). Our third major research objective aims at investigating if goal-setting is an underlying mechanism of achievements. To sum up our research objectives, first, our study aims to investigate motivational benefits of digital achievements as a game design element in gamification with focus on performance and motivation. We deploy a single-element approach to counter confoundation and within an experimental setting for the sake of high controllable requirements and low context-dependency. Second, we aim at investigating design-related aspects of achievements in the form of quantity and difficulty of achievements. Third, we test the proposition of goal-setting being an underlying psychological mechanism of motivational benefits of achievements.

4. Hypotheses

In the following we will deduce all hypotheses in detail. To facilitate the structure and contribution of our hypotheses, we propose three research questions (RQ). Every hypothesis is assigned to one or multiple research questions to contribute to the corresponding research objectives.

RQ1. Are achievements effective in enhancing motivation and performance?

H1.1 to H1.4 address the impact of achievements on motivation and performance. We argue that digital achievements in the form of badges, trophies, etc. directly provide a goal-setting function. Relying on Hamari and Eranti (2011), achievements provide a completion logic, defining the requirements of actions to complete the achievement. In other words, achievements provide clear goals and formulate direct behavioral instructions what an individual has to perform to earn the reward of the specific achievement. For example, in an idea generation task, the completion logic of an achievement may formulate "generate 20 ideas within the first 5 min" (for more examples of achievements see Fig. 1). The phrasing of this achievement may also be exactly formulated as a direct goal. Therefore, an achievement system can be seen as a collection of sub-goal formulations with rewards for completion (usually the award of an achievement-related badge). Many authors have attributed achievements with goal-setting functions (Antin & Churchill, 2011; Fitz-Walter, Tjondronegoro, & Wyeth, 2011; Montola et al., 2009; Sailer et al., 2013). In line with this assumption, we expect similar effects for achievements as for setting specific goals. Thus, given the theoretical and empirical background of the goal-setting theory, we conclude that achievements are also beneficial for motivation and performance.

H1.1. Achievements improve motivation compared to no achievements.

H1.2. Achievements improve performance compared to no achievements.

Following the preceding argumentation, one achievement may not differ much from a set goal. However, achievements are usually integrated into a whole system with a comprehensive set of many sub-goals. The range of the quantity of achievements between systems differs vastly, usually within a two-digit range (de Marcos et al., 2014; Denny, 2013; Hanus & Fox, 2015). We argue that achievement systems contain a lot of information that users have to process and understand. Thus, the full beneficial effects of achievements should not appear instantly, rather require a certain adaption time. We hypothesize that the improvement of motivation and performance due to achievements increase over time.

H1.3. The motivation improvement due to achievements increases over time.

H1.4. The performance improvement due to achievements increases over time.

RQ2. How do achievements need to be designed regarding quantity and difficulty to be effective?

H2.1 to H2.5 address the influence of the achievements' quantity and difficulty on motivation and performance. Beginning with quantity of achievements, we want to emphasize the assumption again that achievements represent sub-goals. In 2009, Ordóñez et al. pointed out the risk of overusing goals and goal-setting. Individuals presented with multiple goals are prone to concentrate only on one goal (Shah, Friedman, & Kruglanski, 2002). Furthermore, some types of goals are more likely to be ignored than others. For example, easily achievable and measurable goals (such as quantity-related) may be given more attention compared to other goals (such as quality-related). These detrimental effects of overuse (i.e., high quantity) may be transferable to achievements as well.

Additionally, as pointed out earlier, fulfilling the completion logic of achievements usually grants a reward. Referencing utility theories, such as the prospect theory (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981), we argue that accumulating achievement rewards results in a reduced value/utility of further rewards. Following the prospect theory, the effort-utility-ratio decreases the more an individual accumulates achievement rewards. Hence, individuals are less motivated to unlock more achievements when they already accumulated a certain amount. In turn, a low quantity of achievements should lead to higher motivation compared to a high quantity, given an integrative achievement difficulty (i.e., the most difficult achievements are identical, irrespective of quantity). Please note that the corresponding performance hypothesis was omitted due to reasons of experimental design (for details see Section 5).

H2.1. A low quantity of achievements leads to higher motivation than a high quantity of achievements.

Further, we assume that quantity directly influences the perception of achievement difficulty. Considering different quantities of achievements from a design point of view, one may argue that a high amount of achievements leads to an inevitable scaling of achievements. Referring to the completion logic of achievements (Hamari & Eranti, 2011), aspects and dimensions may hit a limit regarding latitude, especially in simple tasks. For example, in an idea generation task, certain behavior and performance indicators may be addressed by achievements such as quality, quantity, speed, or accuracy. To reach a high amount of achievements, most achievement systems resort to breaking down a final goal (i.e., a certain amount of quantity) into several sub-goals. Hamari and Eranti (2011) also refer to this as a level hierarchy of achievements where several achievements have a similar completion

Fig. 1. Sample achievements deployed in Experiment 1 (left), Experiment 2 (center), and Experiment 3 (right).

logic (e.g., Achievement A reads "generate 10 ideas", Achievement B reads "generate 20 ideas", etc.). We argue that a high amount of achievements within a system will most certainly include hierarchy based achievements. Low-hierarchy achievements are easier to fulfill which will lead to an early accumulation of completed achievements compared to a system with only the final goal formulated as an achievement. Therefore, we hypothesize that users will perceive achievement difficulty as lower in a system with a high quantity of hierarchy-oriented achievements compared to a system with a low quantity of achievements (given the final goal is identical).

H2.2. A low quantity of achievements leads to higher perceived achievement difficulty than a high quantity of achievements.

The relationship between goal difficulty and motivational benefits as well as performance has been a very robust effect over decades of empirical research. Reviews and meta-analyses (e.g., Locke, 1968; Mento, Steel, & Karren, 1987) as well as numerous studies (e.g., Bassett, 1979; Latham, Mitchell, & Dossett, 1978; Mento, Cartledge, & Locke, 1980; Yukl & Latham, 1978) have shown a positive linear relationship between goal difficulty and performance. Locke and Latham (1990) report goal difficulty effect sizes (*d*) within the range of 0.52–0.82. Goal difficulty represents a crucial core aspect of goal setting theory. Locke and Latham (2005) highlight that goals motivate people by regulating the intensity of effort in line with the difficulty. Given our assumption, that digital achievements provide a goal-setting function in line with the goal-setting theory, we hypothesize that difficult achievements lead to higher motivation and higher performance than easy achievements.

H2.3. Difficult achievements lead to higher performance than easy achievements.

H2.4. Difficult achievements lead to higher motivation than easy achievements.

The differences in performance, emerging under varied goal conditions, indicate an individual adjustment of performance to the goal difficulty (Locke, 1996). However, this adjustment of behavior does not necessarily lead to an increase in performance (Hollenbeck & Klein, 1987; Locke, 1996). Whereas difficult goals may raise performance, low goals mainly result in sustainable lower performance (Locke & Latham, 1990, 2002). Individuals perceive goals as a type of performance threshold, determining the required performance level. In consequence, they tend to adapt their performance to the difficulty of the goal. With regard to easy goals, individuals' adapted performance may not correspond to their actual abilities (Latham, 2015; Latham & Locke, 1991). We assume that these mechanics also apply to the difficulty level of digital achievements, namely, that participants adjust their performance towards the highest achievement, which we call *achievement threshold*.

H2.5. Participants adjust their performance to the achievement threshold.

RQ3. Do achievements share similar mechanics with classical goal-setting?

H2.5 and H3.2 address the direct comparison between achievements and goal-setting. Within the reasoning of all previous hypotheses, we highlight the comparability of digital achievements to goal-setting. In our following last two hypotheses, we focus on a direct comparison between those two approaches. We hypothesize that, in direct comparison, both approaches produce similar results. In detail, setting a specific and high goal on the one hand, and deploying digital achievements on the other, both will increase motivation and performance compared to the absence of both approaches.

H3.1. Achievements improve motivation to the same degree as classical goal-setting.

H3.2. Achievements improve performance to the same degree as classical goal-setting.

For an overview of all hypotheses see Table 1. It also depicts the assignment of hypotheses to the research questions as well as to the experiments. In the next sections, we will present the latter in detail.

5. Experiment 1

The first experiment focuses the relationship between the quantity

Table 1
Overview of hypotheses in conjunction with corresponding experiments, addressed research questions, and results.

No.	Hypothesis	Experiment	RQ	Confirmed
H1.1.	Achievements improve motivation compared to no achievements.	1–3	1	partially
H1.2.	Achievements improve performance compared to no achievements.	2 & 3	1	yes
H1.3.	The motivation improvement due to achievements increases over time.	3	1 & 3	no
H1.4.	The performance improvement due to achievements increases over time.	3	1 & 3	yes
H2.1.	A low quantity of achievements leads to higher motivation than a high quantity of achievements.	1	2	partially
H2.2.	A low quantity of achievements leads to higher perceived achievement difficulty than a high quantity of achievements.	1	2 & 3	yes
H2.3.	Difficult achievements lead to higher performance than easy achievements.	2	2 & 3	yes
H2.4.	Difficult achievements lead to higher motivation than easy achievements.	2	2 & 3	no
H2.5.	Participants adjust their performance to the achievement threshold.	2	2 & 3	partially
H3.1.	Achievements improve motivation to the same degree as classical goal-setting.	3	3	no
H3.2.	Achievements improve performance to the same degree as classical goal-setting.	3	3	partially

of achievements and motivation. As we chose to measure motivation through persistence, time on task varies between subjects. Therefore, performance indicators are problematic to interpret; hence we decided not to analyze performance in this experiment. Experiment 1 was designed to test Hypotheses H1.1, H2.1, and H2.2 (see Table 1).

5.1. Method

5.1.1. Design

We deployed a one-factor between-subjects design. The factor represented the quantity of achievements on three levels: no achievements versus low quantity versus high quantity. Dependent variables were motivation, measured by persistence operationalized via self-determined time on task in an online Stroop test (MacLeod, 1991; Stroop, 1935), as well as achievement difficulty. As control variables, we included achievement motive, task complexity, and achievement commitment. The first has already shown to influence persistence (e.g., Feather, 1961; Smith, 1964), both latter are essential moderators of goal-setting (Locke & Latham, 2002). Thus it was advisable to control for potential third variable effects.

5.1.2. Participants

Participants were recruited online as well as via bulletin boards. Out of 120 participants, 93 provided complete data, and, $N = 87$ were included in the data analyses (see Section 5.2). Of the participants, 62 (71.26%) were female. The average age was 31 ($M = 30.90$, $SD = 15.43$). Students ($n = 54$, 62.07%) were eligible for course credit as a task-noncontingent reward. Latter has shown to not influence intrinsic motivation because task-noncontingent rewards do not require to complete the task or perform specifically well (Deci, Koestner, & Ryan, 1999). No monetary incentive was offered for participation.

5.1.3. Materials

As the main task, we deployed the Stroop test, also known as the Color-Word Interference Test. Developed by Stroop (1935), it suits to measure executive functions, selective attention, and cognitive flexibility (Homack & Riccio, 2004). In this test (for an overview of different versions, see MacLeod, 1991), participants are requested to name the color of the ink in which a color name is presented. In some trials, the ink color presented conflicts with the printed name (e.g., the word *RED* printed in green letters; incongruent); in other trials, colors are not in conflict with the color name (e.g., *RED* printed in red letters; congruent). Control trials include only a colored square without a word. We decided to use the Stroop test, as it is easy to learn without previous knowledge, its structure provides a suitable base for designing achievements, and participating becomes monotonous and tedious relatively fast.

To assess motivation, we measured persistence with time-on-task, as it is widely used in motivation literature (Deci, 1971; Deci et al., 1999; Elliot & Harackiewicz, 1994). We gave participants a free choice during the Stroop test to quit the task whenever they choose to and switch to an alternative task. The secondary task was a simple computerized math task as presented in Turner et al. (1986) or Girdler, Turner, Sherwood, and Light (1990). The results of the math task were of no relevance to our study. It served the sole purpose to occupy all participants with an identical participation time duration in sum. With this approach, we intended to avoid participants receiving course credit compensation despite aborting the Stroop test immediately. Therefore, we introduced the second task to keep the overall participation time identical, only the proportion of task one and task two varied due to participants' free choice.

Achievement difficulty was assessed with an adapted item from a questionnaire by Lee and Bobko (1992) measuring subjective goal difficulty, using a five-point Likert scale. We only deployed item number five of the self-reference goal difficulty scale, because the other four were not transferable to the achievement context ("How difficult were the achievements you had to attain for this experiment?"; 1 = *not at all difficult* to 5 = *extremely difficult*).

We adapted the four items for achievement commitment out of an inventory for measuring goal commitment by Hollenbeck, Klein, O'Leary, and Wright (1989) (e.g., "It's hard to take these achievements seriously"; "Quite frankly, I don't care if I attain these achievements or not"; five-point Likert scale; 1 = *strongly disagree* to 5 = *strongly agree*), yielding a reliability of $\alpha = .79$.

Achievement motive was measured with a subscale of the Achievement Motives Scale (AMS) by Lang and Fries (2006), namely the Hope for Success Scale ($\alpha = .78$). It includes five items with a four-point Likert scale (e.g., "I like situations, in which I can find out how capable I am"; "I am attracted by tasks, in which I can test my abilities"; 1 = *strongly disagree* to 5 = *strongly agree*).

For task complexity, we used an item of a questionnaire by Maynard and Hakel (1997) deploying a seven-point Likert scale ("I found this to be a complex task"; 1 = *totally disagree* to 7 = *totally agree*).

We designed the achievement system in reference to the recommendations of Hamari and Eranti (2011). Accordingly, each achievement was designed to comprise a signifier (i.e., name, visual, and description), completion logic (i.e., pre-requirements and conditions to complete an achievement), and a reward (i.e., badge as a representation of the attained achievement). An example of achievements, used in the experiments, can be reviewed in Fig. 1. The condition with a high quantity of achievements included 30 achievements, incentivizing quantity (number of finished trials), quality (number of correct trials), and accuracy (number of subsequent correct trials). Each of the three categories was divided into ten smaller achievement levels (e.g., "finish 10 trials", "finish 20 trials", "finish 50 trials"; see Hamari & Eranti, 2011 for levels in achievements). The low quantity condition only included three achievements. They incentivized the same three aspects and were identical to the highest level achievements in the other condition (e.g., "finish 700 trials"). This approach was to ensure an equal difficulty at the end of the achievement spectrum.

5.1.4. Procedure

The experiment was designed and programmed with Inquisit Lab (Version 4.0.10.0) by Millisecond Software. It was hosted on a dedicated private server using Inquisit Web Player (Version 4.0.10.0) for individuals to participate. The player did not support the use of mobile devices.

After providing consent and demographics, participants were instructed to work on two tasks within the next 20 min. They had free choice of when to switch to the second task. The subject of the latter was not conveyed to avoid triggering preferences in advance. A switch back to the first task was not possible. After that, instruction for the Stroop test was provided. Colors red, green, blue, and black were mapped to the keys D, F, J, and K, respectively. Participants were instructed to press the key associated with the color shown on the screen. Before the main task, achievements were introduced along with all potential achievements to unlock (except control group).

Achievements were always described as an optional bonus and supplemental to the main task. We deliberately avoided wording which possibly would have had a resemblance to goal-setting in any way (e.g., "your goal will be to unlock achievements"). During the Stroop test, participants received feedback for wrong answers (red cross for 400 ms) and no feedback for right answers. Keyboard commands were always visible (i.e., key mapping, switch command). Whenever the completion-logic of an achievement was fulfilled, participants received a notification in the right lower corner of the screen immediately showing the unlocked achievement. Every 80 trials, a pause interrupted the task showing an overview of all achievements (not in the control group). All accomplished achievements were shown multicolored, the rest of still unaccomplished achievements were shown in grey. After participants chose to switch the task (or after 20 min of working on the Stroop test),

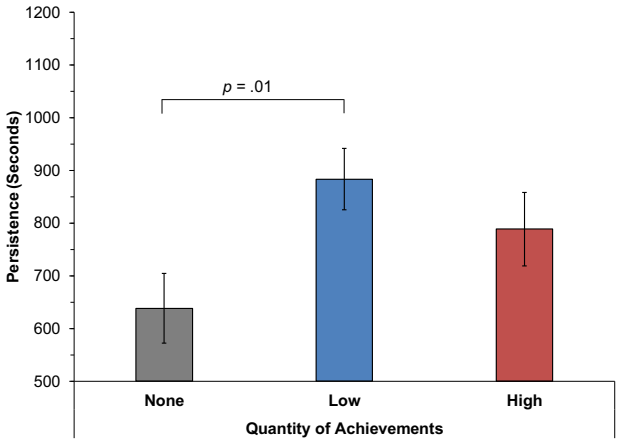

Fig. 2. Means of persistence (seconds) of all respective experimental groups in Experiment 1. Bonferroni-corrected pairwise comparisons revealed a significant difference between the group with a low quantity of achievements and the group with no achievements. Error bars denote one standard error around the respective mean of each group.

a questionnaire with the self-report variables (see Section 5.1.3) was deployed. All measures referred to the Stroop test. After the 20 min, participants received debriefing, thanks, and the experiment ended.

5.2. Results

Out of 93 participants who provided complete data, six were excluded due to non-reasonable response patterns (e.g., random responses, extreme long response times). We will not report analyses of control variables, as results provided no relevant insights with regard to the hypotheses. For recall, Hypotheses H1.1, H2.1, and H2.2 were to be tested (see Table 1). Level of significance of all conducted statistical tests was at an alpha level of .05. All directional hypotheses were one-tail tested. This proceeding applies to the analyses of all experiments.

The means of the groups were $m = 638.31$ s ($n = 29$, $SD = 356.42$) for the no achievements group, $m = 883.37$ s ($n = 27$, $SD = 301.64$) for the low quantity achievements group, and $m = 788.52$ s ($n = 31$, $SD = 391.43$) for the high quantity achievements group (see Fig. 2).

To test for Hypotheses H1.1 and H2.1, we conducted a one-way analysis of variance (ANOVA) with Bonferroni-corrected post hoc multiple comparisons. Results of the ANOVA showed a significant main effect with medium effect size for the experimental condition factor ($F(2,84) = 3.43$, $p = .037$, $\eta_p^2 = .08$). Post hoc multiple comparisons showed a higher persistence for the low quantity achievements group compared to the no achievements group ($MD = 245.06$, 95% CI [13.87, 476.25], $p = .01$). However, neither did the high quantity achievements group show a significantly higher persistence than the no achievements group ($MD = 150.21$, 95% CI [-73.12, 373.54], $p = .16$), nor did the low quantity achievements group compared to the high quantity achievements group ($MD = 94.85$, 95% CI [-132.71, 322.42], $p = .47$). Effect sizes for mean differences were small ($d_{low-high} = 0.27$), medium ($d_{no-high} = 0.40$) and large ($d_{no-low} = 0.74$). Results confirm both Hypotheses H1.1 and H2.1 partially because only low quantity achievements showed higher persistence than no achievements (H1.1). Also, though the achievement groups did not differ in direct comparison, the low quantity group showed higher persistence than the control group, whereas the high quantity group did not (H2.1).

To test for Hypothesis H2.2, we conducted a non-parametric Mann-Whitney U Test (due to a violation of normality and violation of homogeneity of variance of the sample) to test for mean differences of perceived achievement difficulty between the two experimental groups of low and high achievement quantity. The mean for perceived achievement difficulty was $m = 3.48$ ($SD = 1.22$) in the low quantity group and $m = 2.29$ ($SD = 0.68$) in the high quantity group (see Fig. 3). The difference of the means ($MD = 1.19$) showed to be significant ($U = 191{,}50$ $p < .001$) with a large effect size ($d = 1.14$). Thus, the results support Hypothesis H2.2.

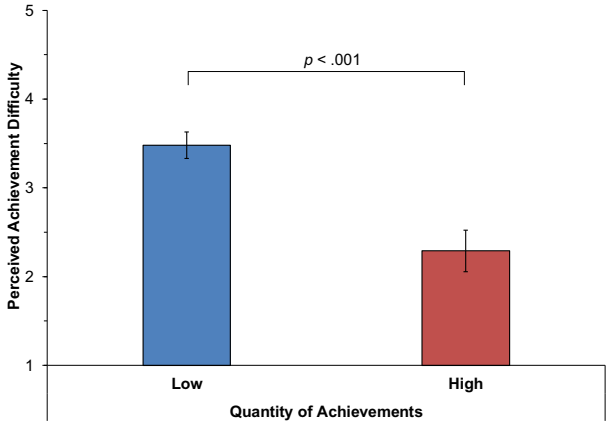

Fig. 3. Means of perceived achievement difficulty (five-point scale) in Experiment 1. Results of a U test for mean differences revealed that participants in the low quantity group perceived the deployed achievements as harder than the high quantity group. Error bars denote one standard error around the mean of each group.

5.3. Discussion

Partially confirming Hypothesis H1.1 and H2.1, we conclude that achievements are effective in promoting motivation measured by persistence; however, the effectiveness of achievements depends on the quantity of deployed achievements. A high quantity of achievements appears to be non-effective or at least less effective compared to a low quantity of achievements. An explanation may be found in the confirmation of Hypothesis H2.2. Evidence showed that a high quantity of achievements leads to a perception of lower difficulty compared to a low quantity. This perception arises probably due to the existence of lower achievement levels in high quantity systems, leading individuals to unlock several achievements relatively early throughout the task. This early unlocking leads to the impression that achievements are easy to achieve compared to a low quantity system, where achievements incentivize only the *end-goal*. This finding is in accordance to the goal-setting theory which postulates that difficult goals are more beneficial for effort and persistence than easy goals (Latham, 2015).

6. Experiment 2

The second experiment focuses the relationship between the difficulty of achievements and motivation, as well as performance. It was designed to test Hypotheses H1.1, H1.2, H2.3, H2.4, and H2.5 (see Table 1).

6.1. Method

6.1.1. Design

In this experiment, we deployed a 3 × 2 two-factor mixed design. The first factor (between-subjects) represented the difficulty of achievements on three levels: no achievements versus easy achievements versus difficult achievements. The second factor (within-subjects) represented two tasks, the first task with achievements absent, the second task with achievements present. This approach allowed us to obtain a reference measure for every participant and investigate the precise performance adjustment due to the introduction of achievements. As a task, we chose to deploy an online picture annotation task, similar to the task used in other gamification studies (Mekler,

Brühlmann, Tuch, & Opwis, 2015; Mekler et al., 2013). Dependent variables were performance and self-reported motivation. As control variables, we included achievement commitment as well as subjective achievement difficulty, the latter as a manipulation check to assess if participants experienced the difficult achievements as more difficult than the easy achievements.

6.1.2. Participants

We recruited participants both online as well as via bulletin boards. Out of the 115 participants who attended the online experiment $N = 99$ participants provided complete data. The average age was 26 ($M = 25.97$, $SD = 9.33$). The majority were female ($n = 67$, 67.68%). Students ($n = 71$, 71.72%) were eligible for course credit. No monetary incentive was offered for participation.

6.1.3. Materials

The main task was split into two parts to facilitate the within-subject factor. Both parts comprised a picture annotation task. Participants were asked to generate as many keywords as possible for presented photographs and pictures. Latter varied in content, such as landscapes, animals, people, and events. Eighteen pictures were presented in total, nine for each task. Each picture was presented for 90 s. We instructed participants not just to give a verbal description of the picture but rather generate single keywords with associations which come to mind, comparable with generating hashtags in social media. For example, a picture of a beach may elicit keywords such as *relaxing, holiday, hot, swimming, sunburn*, which count as five annotations for the respective picture. We decided to use a picture annotation task as it showed to be suitable to measure performance in an experimental gamification environment (Goh & Lee, 2011; Mekler et al., 2013). Due to its low complexity and neutral context, it is less interference-prone with regard to factors such as cognitive ability or personal preference.

We measured performance as the overall number of picture annotations, separately for both halves of the task. For comparability reasons, we used the standardized value of annotations per minute in data analysis.

Motivation was measured with the subscale interest/enjoyment of the Intrinsic Motivation Inventory (IMI) by Ryan (1982) using a seven-point scale (1 = *strongly disagree* to 7 = *strongly agree*). This subscale is considered the self-report measure of intrinsic motivation (Deci, Eghrari, Patrick, & Leone, 1994; Popović, Kostić, Rodić, & Konstantinović, 2014). McAuley, Duncan, and Tammen (1989) found strong support for its validity. It contains seven items (e.g., “I thought this activity was quite enjoyable“; “I would describe this activity as very interesting“) In the present sample Cronbach's α was .92.

We assessed achievement commitment with the inventory for goal commitment by Hollenbeck et al. (1989), and adapted it slightly to match the achievement context ($\alpha = .78$).

To check if the manipulation was successful, we deployed two items to ask participants how difficult they perceived the achievements on a five-point scale (“It was difficult to attain the achievements”; “To attain the achievements, a high performance was necessary“; 1 = *strongly disagree* to 5 = *strongly agree*).

The basic design for achievements was the same as for the first experiment, following the recommendations of Hamari and Eranti (2011) (see Section 5.1.3 and Fig. 1). Achievements could be unlocked for quantity only because there were no regulations of correct and false answers (however, nonsensical answers were removed during data processing afterward). We deployed achievements with a level hierarchy (see Section 4) subdivided into labels with *bronze*, *silver*, and *gold*. We also added three—as we call it—*meta achievements* which were unlocked for attaining all bronze achievements, respectively silver and gold. Each achievement referred to three pictures resulting in a total number of twelve achievements to obtain for each participant. We aligned the difficulty of the achievements with the results of a prestudy with a similar task within a gamification environment. In the prestudy, participants in the control group (no achievements) generated 21 keywords per picture on average. Around that, we built the achievement system. To attain a bronze achievement in the difficult condition, participants needed to generate 20 keywords per picture, respectively 30 for silver and 40 for gold. In the easy condition, the scaling was 5, 10 and 15, respectively. Please note that we report the standardized value keywords/annotations per minute for reasons of comparability. Participants were always given the exact absolute number of annotations when referring to achievements or when given feedback (for example, to generate 90 annotations during the first three pictures to attain a silver achievement in the difficult condition).

6.1.4. Procedure

Designing and programming were identical to the first experiment. After giving consent regarding data privacy and voluntariness, participants were introduced to the picture annotation task. They were instructed to type as many annotations as possible for the presented pictures. There were no limitations concerning content or annotation type (such as nouns, adjectives, verbs). After two example trials, participants started with the first task, which was identical for all participants and comprised nine pictures with a duration of 90 s. After each picture, participants could choose when to continue with the next picture, so a pause between pictures was possible. After the first task, achievements were introduced for the respective groups with detailed information on how to attain them. Achievement awarding, descriptions, and overview pages had the same style as in the first experiment (see Section 5.1.4). The second task had the same structure as the first task. After every third picture an overview of achievements was shown (those still to be awarded were greyed out). After the second task, participants filled out the questionnaires (see Section 6.1.3). A debriefing together with thanks ended the experiment after around 45 min.

6.2. Results

Out of 99 participants, who provided complete data, all were included in data analysis. The manipulation check showed to be successful. Participants in the group of difficult achievements perceived the achievement as more difficult ($m = 4.21$, $SD = 0.89$) than in the easy achievements group ($m = 2.41$, $SD = 1.09$). This difference reached statistical significance ($MD = 1.80$, $t = 7.22$, 95% CI [1.30, 2.30], $p < .001$) with a large effect size ($d = 1.81$). For recall, Hypotheses H1.1, H1.2, H2.3, H2.4, and H2.5 were to be tested (see Table 1).

The means for generated annotations per minute in the first task (without achievements) were $m_{control} = 8.18$ ($SD = 3.08$), $m_{low} = 8.02$ ($SD = 2.70$), and $m_{high} = 8.25$ ($SD = 4.09$); in the second task (with achievements) means were $m_{control} = 8.07$ ($SD = 3.19$), $m_{low} = 9.97$ ($SD = 2.42$), and $m_{high} = 12.22$ ($SD = 5.29$). For an overview, see Fig. 4. The overall means for motivation were $m_{control} = 4.46$ ($SD = 1.24$), $m_{low} = 4.33$ ($SD = 1.32$), and $m_{high} = 4.45$ ($SD = 1.34$).

To test for Hypotheses H1.1, H1.2, H2.3, and H2.4 we conducted a one-way ANOVA with post hoc multiple comparisons for the second task, both for performance as well as for motivation. Homogeneity of variance was violated for performance; thus we Welch-corrected the F-value and Games-Howell-corrected the post hoc comparisons. For motivation, no effects were revealed ($F(2,96) = 0.10$, $p = .90$, $\eta_p^2 < .01$). Thus, Hypotheses H1.1 and H2.4 have to be rejected. For performance, results of the ANOVA showed a significant main effect for the group factor ($F(2,58.49) = 8.13$, $p = .001$) with a large effect size ($\eta_p^2 = .22$). The post hoc multiple comparisons revealed a higher performance for the high difficulty group compared to the control group ($MD = 4.15$, 95% CI [1.15, 6.79], $p < .001$). Also, the low difficulty group outperformed the control group ($MD = 1.90$, 95% CI [0.26, 3.54], $p = .01$). The fact that both experimental groups outperformed the

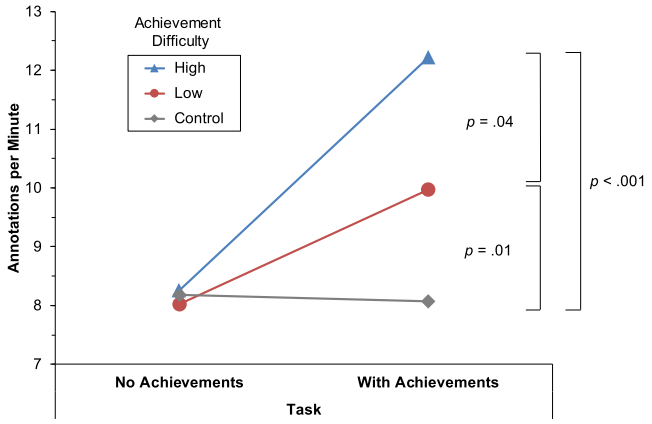

Fig. 4. Interaction between task and achievement difficulty in Experiment 2. In the first task, no group was presented with achievements. In the control condition, achievements were absent in both tasks. Each picture lasted for 90 s. Bonferroni-corrected pairwise comparisons showed that all groups differed from each other in the second task.

control group fully supports Hypothesis H1.2. Furthermore, the two experimental groups also differ from each other; the high difficulty group outperformed the low difficulty group (MD = 2.25, 95% CI [0.27, 4.78], p = .04). Thus, Hypothesis H2.3 is also confirmed.

To test Hypothesis H2.5, we first computed a general linear model to test the within-subjects effects of the interaction of task and group. The results show that the interaction is statistically significant (F(2, 68.58) = 38.40, p < .001, η_p^2 = .53). This interaction means that the difference in performance between the first and the second task depends on the group, namely the difficulty of the achievements (see Fig. 4). To get further insights into the threshold hypothesis, we conducted one-sample t-tests to test the respective group mean of the second task against the given achievement thresholds. We hypothesize that, on average, participants align their performance towards the highest achievement. Results show that participants' mean performance (annotations per picture) in the easy achievements group (m = 14.96, SD = 3.63) was statistically indifferent from the highest achievement of 15 annotations per picture (MD = -0.04, t = -0.07, 95% CI [-1.33, 1.24], p = .95), whereas all other achievements where significantly lower than the participants' performance. These results indicate towards an alignment of the performance towards the highest achievement. For the difficult achievements group, results show that participants' mean performance (m = 18.34, SD = 7.93) was statistically indifferent from the lowest achievement of 20 annotations per picture (MD = -1.66, t = -1.17, 95% CI [-4.57, 1.25], p = .25), whereas all other achievements where significantly higher than the participants' performance. This indicates towards an alignment of the performance towards the lowest achievement.

Taking together, results indicate that participants show a performance increase due to the introduction of achievements in the second task, whereas the control group's performance remains unchanged. The threshold Hypothesis H2.5, however, was only partially confirmed. Whereas participants in the low difficulty group show an alignment towards the highest achievement level, participants in the high difficulty group show an alignment towards the lowest achievement level.

6.3. Discussion

In conclusion, Hypotheses H1.2 and H2.3 were confirmed, H1.1 and H2.4 could not be confirmed, and H2.5 was partially confirmed. We derive from the results that achievements are effective in promoting performance. Both groups, in which achievements were present, outperformed the control group without achievements. Additionally, the difficulty of the deployed achievements plays an essential role. Difficult achievements resulted in higher performance than easy achievements. These results are directly in line with the goal setting theory (Locke et al., 1981) and support the assumption that achievements adhere to a goal-setting function where difficult achievements lead to higher performance similar to classic goal formulation. However, no effects of achievements could be revealed with regard to motivation. Participants did not differ in their reported motivation. This result is in contradiction to the first experiment. On the one hand, one may conclude that achievements are not able to enhance motivation, on the other hand, in the first experiment, motivation was operationalized by persistence whereas in the second experiment by self-report. We will further discuss this issue in the general discussion (see Section 8).

Furthermore, results showed that participants directly adapted their performance after the introduction of achievements. In the easy achievements group, the group mean was almost identical to the highest achievement attainable. This fact supports the assumption that participants try to attain the highest achievement, but may not try to exceed that specific achievement any further. This threshold assumption also may explain why participants in the difficult achievements group outperformed the easy achievements group. They were not affected by the threshold, because there still were achievements open to attain. A reason, why they settled around the lowest achievement and did not perform better than that, may be found in the consideration that they already reached their performance maximum. A mean of 4.21 on the achievement difficulty scale and the fact that 42% of the group responded with the maximum of 5 points show strong support that the achievements in the high difficulty condition were too difficult to attain. Therefore, the lowest achievement was the most realistic option to choose from; nevertheless, it was still higher than the highest achievement of the low difficulty group, which presumably led to a better performance.

7. Experiment 3

The third experiment focuses the comparison of the impact of achievements on motivation and performance with classical goal setting. It also allows for a second ascertainment to take a time-dependent effect into account. Experiment 3 was designed to test Hypotheses H1.1, H1.2, H1.3, H1.4, H3.1, and H3.2 (see Table 1).

7.1. Method

7.1.1. Design

For this experiment, we deployed a one-factor between-subjects design with a one week delayed second measurement. The experimental factor comprised three levels: no achievements versus goal-setting instructions versus achievements. We chose idea generation for the task at hand because idea generation tasks have been successfully deployed in goal setting research for decades (e.g., Jung, Schneider, & Valacich, 2010; Latham & Locke, 1979; Latham & Saari, 1979; Locke, 1966). Due to the length of the task, the multi-session nature, and technical reasons, we deployed the experiment in a computer-aided laboratory environment. For dependent variables, we measured motivation and performance. Control variables were goal specificity and goal difficulty to assure for an adequate goal core as well as feedback notice, goal commitment, self-efficacy, task complexity, and goal importance to control for the essential moderators in goal-setting and the high-performance cycle (Locke & Latham, 2002).

7.1.2. Participants

For participant acquisition, we recruited on bulletin boards as well as online on social media platforms. In sum, 65 individuals participated in the experiment. We excluded four participants due to their absence in Session 2. Furthermore, two more participants were excluded, due to erroneous subject matching. The final sample comprised N = 59

participants. On average, participants were 22 years old (M = 21.76, SD = 3.36). The majority was female (n = 44, 75.58%). The sample comprised solely of students, most with a psychology major (n = 55, 93.22%). All were eligible for course credit. No other incentives were offered.

7.1.3. Materials

The main task—an idea generation task—was split into two sessions (one week apart) with two subtasks each, which resulted in four separate subtasks with individual idea generation exercises. We used types of idea generation tasks which are established in the literature such as generating improvement ideas for a specific topic or problem, generating uses for an object, and the *thumbs problem* (e.g., Diehl & Stroebe, 1991; Gallupe et al., 1992; Locke, Frederick, Lee, & Bobko, 1984; Paulus & Dzindolet, 1993). More precisely, we asked participants to generate ideas to enhance the local student canteen, to generate ideas to enhance the local tourism situation, to generate uses for a kitchen knife, and to generate problems as well as opportunities, if humans were growing a second thumb on each hand.

We measured performance as the sum of ideas generated. Motivation was measured identical to Experiment 2 with the subscale interest/enjoyment of the IMI by Ryan (1982). Cronbach's α was .94 in this sample.

We measured goal specificity with three self-formulated items on a five-point scale (1 = *strongly disagree* to 5 = *strongly agree*; e.g., "The goal was specific", "I knew what to do to attain the goal"; α = .61) and goal difficulty with four items adapted from Lee and Bobko (1992) such as "The goal I was given of 45 ideas per task required:" (five-point scale from 1 = *very little attention and effort* to 5 = *as much attention and effort I can give*; verbal anchors varied for each item; α = .89). We also asked participants, if they noticed the feedback presented (i.e., the current number of ideas generated; dichotomous) and whether they used this feedback to check their status quo of ideas (five-point scale from 1 = *never* to 5 = *constantly*). Goal commitment was measured with a five-item scale from Klein, Wesson, Hollenbeck, Wright, and DeShon (2001; five-point scale from 1 = *strongly disagree* to 5 = *strongly agree*; e.g., "I am strongly committed to pursuing this goal", "I think this is a good goal to shoot for"; α = .79). To measure task-related self-efficacy, we created one item in concordance with Bandura (2006; i.e., "Now, that you were introduced to the task, how confident are you to accomplish this type of task?"; ten-point scale from 1 = *cannot do at all* to 10 = *highly certain can do*). We collected data for task complexity using the Subjective Task Complexity Scale by Maynard and Hakel (1997; seven-point scale from 1 = *totally disagree* to 7 = *totally agree*; e.g., "This task required quite a bit of thought and problem solving", "I found this to be a complex task"; α = .82). Goal importance was measured with a single item on a nine-point scale (1 = *not at all important* to 9 = *very important*) asking "How important was it for you to attain the given goal?" (Martin & Gill, 1995). Please note that all presented goal related items were adapted for the achievements condition to state *achievements* instead of *goal*, as participants were not provided with goal-setting.

Design of achievements followed a similar pattern as in Experiment 1 and Experiment 2. We also used a hierarchy structure as well as meta-achievements (see 5.1.3). For performance, we awarded bronze for 10, silver for 25, and gold for 45 ideas per single task (respectively 0.5, 1.25, and 2.25 ideas per minute). These values were based upon a designated prestudy in which participants generated 22 ideas per task on average (N = 79, M = 22.47, SD = 11.65). Due to the extended length of one session (50–60 min, including instructions, tasks, questionnaires, and debriefing), as well as the dual-session nature, we extended the achievement system by other categories. These were for example achievements for speed, collecting a certain amount of other achievements (meta achievements), or achievements for completing several stages (e.g., reaching half of the session, filling out all questionnaires). With this, we intended to keep the awareness of achievements on a constant level. In total, participants could attain 45 achievements over the course of the experiment.

7.1.4. Procedure

Designing and programming were identical to the former experiments, with the exception that no Inquisit Web Player was needed because hosting took place in a laboratory environment. After giving consent, the experimenter welcomed participants and asked to generate an individual code to match the two experimental sessions with the pen and paper questionnaire. Afterward, participants were instructed with details about the upcoming task. We gave very detailed instructions on how to formulate ideas (i.e., examples or detailed aspects of previous ideas should be entered as new separate ideas, avoidance of extensive reasoning of ideas, etc.). We found this necessary to ensure adequate comparability as well as appropriate live awarding of achievements. Participants of the respective group were provided with an introduction and descriptions of the achievement system as well as an overview of all attainable achievements in advance. Every idea generation task lasted 20 min. By request (pressing a specific key), participants could review attained and yet receivable achievements (reviewing time was substracted from the total editing time). A review of achievements was also possible between tasks. Self-efficacy was measured before each task. After each task, motivation as well as prior knowledge related to the attended task's topic were measured. Attained achievements of Session 1 were transferred to Session 2. Control variables were all measured at the end of Session 2 via paper-pencil questionnaire. The experiment ended with a debriefing.

7.2. Results

Before data analysis, all reported ideas were manually reviewed for doublets as well as purportless inputs and removed afterward. No further participants were excluded because all provided meaningful responses and showed no indications of not taking the task seriously (i.e., notable outliers for low self-reported motivation, low performance, or unrealistically high response times). For recall, Hypotheses H1.1, H1.2, H1.3, H1.4, H3.1, H3.2 were to be tested (see Table 1).

Means for the respective groups were $m_{control}$ = 1.56 (SD = 0.67), m_{goal} = 1.96 (SD = 0.77) and $m_{achievement}$ = 2.41 (SD = 1.04) ideas per minute (see also Fig. 5). For motivation, means were as follows: $m_{control}$ = 4.37 (SD = 0.80), m_{goal} = 4.31 (SD = 0.90), $m_{achievement}$ = 4.16 (SD = 0.58).

We conducted one-way ANOVAs with post hoc multiple comparisons to test for Hypotheses H1.1, H1.2, H3.1, H3.2 for performance as well as for motivation. The analysis for performance was Welch-corrected as well as Games-Howell-corrected due to a violation of homogeneity of variance. Motivation analysis revealed no effects (F(2,56) = 0.35 p = .71 η_p^2 = .01). Also, no group differences (Bonferroni-corrected) were observed between the groups. Hence, Hypotheses H1.1 and H3.1 have to be rejected. Performance analysis however did reveal a large effect (η_p^2 = .21) for the group factor (F(2,35.93) = 4.71 p = .02). In detail, the achievement group outperformed the control group (MD = 0.85, 95% CI [0.15, 1.55], p = .007) with a large effect size (d = 0.97). However, neither did the goal-setting group outperform the control group (MD = 0.40, 95% CI [-0.15,0.96], p = .10; despite a medium effect size d = 0.56), nor did the achievement group outperform the goal-setting group (MD = 0.45, 95% CI [-0.28,1.17], p = .15, d = 0.49). For an overview see Fig. 5. These results support Hypothesis H1.2 fully and Hypothesis H3.2 partially.

We tested Hypotheses H1.3 and H1.4 with two repeated measures ANOVAs, for performance and motivation respectively, group as the between-subject factor and session as the within-subject factor. If necessary, violations of sphericity were corrected with the Greenhouse-Geisser correction. The analysis for motivation provided a small and marginal significant negative main effect of session (F(1,56) = 3.22,

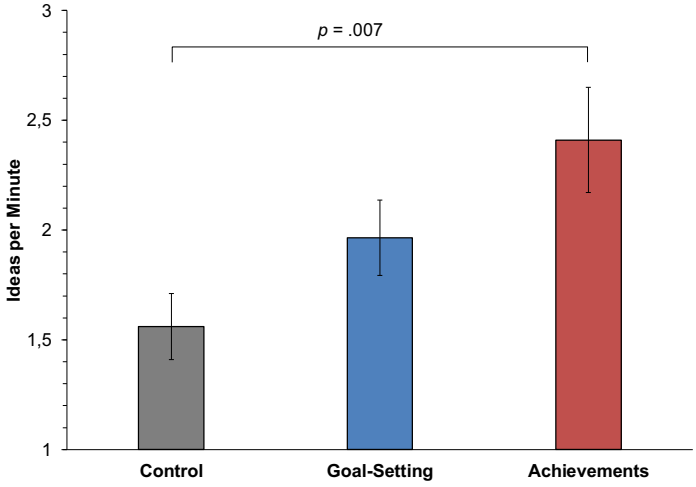

Fig. 5. Means of performance (ideas per minute) of all respective experimental groups in Experiment 3. Bonferroni-corrected pairwise comparisons revealed a significant difference between the group with achievements and the group without achievements (control). Error bars denote one standard error around the respective mean of each group.

$p = .08$. $\eta_p^2 = .05$). This may indicate that participants reported a slightly lower motivation in Session 2 compared to Session 1. However, Bonferroni-corrected pairwise comparisons showed no session effect for the separate groups. An interaction between session and group was also not observed ($F(2,56) = 0.24, p = .79, \eta_p^2 < .01$). Therefore, Hypothesis 1.3 has to be rejected.

Results revealed two effects for performance. The first is a large main effect for session ($F(1,56) = 20.90$, $p < .001$, $\eta_p^2 = .27$), which means participants performed better in the second session compared to the first session (see Fig. 6). The second effect is a medium—however marginal significant—effect for the interaction between session and group ($F(2,56) = 2.60$, $p = .08$, $\eta_p^2 = .09$). Furthermore, Bonferroni-corrected pairwise comparisons revealed that only the achievements group significantly increased their performance between sessions ($MD = 0.71$, 95% CI [0.39,1.04], $p < .001$, $d = 0.59$), whereas the other groups did not, neither the goal-setting group ($MD = 0.28$, 95% CI [-0.03,0.60], $p = .12$, $d = 0.35$), nor the control group ($MD = 0.26$, 95% CI [-0.06,0.57], $p = .16$, $d = 0.37$. Additionally, in the second session, the achievement group outperformed both the goal-setting group ($MD = 0.70$, 95% CI [0.04,1.36], $p = .19$, $d = 0.87$), as well as the control group ($MD = 1.12$, 95% CI [0.46,1.78], $p < .001$, $d = 1.00$). Hypothesis H1.4 can be confirmed.

7.3. Discussion

Concluding hypotheses testing, Hypotheses H1.2 and H1.4 were fully confirmed, Hypothesis H3.2 partially confirmed, whereas Hypotheses H1.1, H1.3, and H3.1 had to be rejected. These results are in line with the first and the second experiment concerning the fact that achievements led to a performance improvement compared to no achievements. We conclude that achievements are effective in increasing performance.

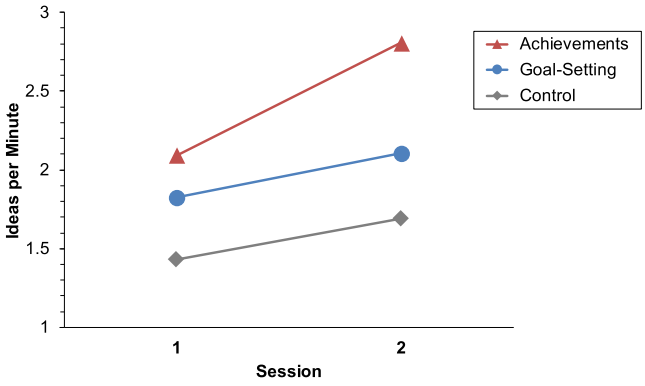

Fig. 6. Interaction between experimental condition and session in Experiment 3. The main effect of session was also statistical significant. The achievements group significantly outperformed all other groups in Session 2. Control: no goals, no achievements; Goal-Setting: only goals; Achievements: only achievements.

Furthermore, no differences in motivation could be observed. These results match the observations of the second experiment, where effects of motivation were also absent, but contradicts the results of the first experiment where effects of motivation were present. However, operationalization differed in the first experiment compared to the other two. We will address further considerations in the general discussion.

The assumption that the outperformance of goal-setting and achievements compared to the control group are similar was only partially confirmed. Although the achievements group outperformed the control group, but not the goal-setting group, the goal-setting group did not outperform the control group (see Fig. 5). However, we argue that, given the effect size of $d = 0.56$ for the mean difference between the goal-setting and the control group, and considering former goal-setting research, which observed mean effect sizes around $d = 0.60$ (Mento et al., 1987), the nonsignificant effect may be attributed to the merely medium power of the third experiment ($1\text{-}\beta = .80$). In combination with the observed results, indicating that the achievement group increased performance over sessions whereas the other groups did not, we conclude that achievements share similar mechanics with classical goal-setting; however, there seem to be additional benefits from achievements which do not apply to goal-setting. We will discuss this topic further in the general discussion.

8. General discussion

The aim of this work was to investigate digital achievements as a game design element of gamification separately and unconfounded from other elements, their impact on motivation and performance in controlled laboratory/online experiments, as well as their similarity and comparability to classical goal-setting. For structuring this section, we will divide the discussion into the devised research questions. We will address theoretical implications, followed by limitations as well as prospects, and close up with a conclusion.

8.1. Theoretical implications

RQ1. Are achievements effective in enhancing motivation and performance?

Addressing the results of motivation first, they showed to be contradictory. In Experiment 1, we found that participants, introduced

with achievements, tend to show higher persistence by voluntarily switching later to a second task while longer attending to the actual first task. We explain this effect with our proposition that achievements provide a goal-setting function. Achievements provide a clear goal in advance what to do to fulfill the requirements of the achievement. This goal-setting directly affects persistence (Locke & Latham, 2002). On the other hand, in the second and third experiment, self-reported motivation did not differ between groups presented with achievements and other groups without achievements. This finding is in line with other experiments on gamification. For example, Mekler et al. (2013) as well as Mekler et al. (2015) also found no effects on self-reported motivation. The authors also measured motivation with the IMI (Ryan, 1982) and argue that game design elements function as extrinsic incentives, only effective in promoting performance rather than motivation (Mekler et al., 2015). We argue that the contradiction in our findings of motivation is attributed to the operationalization used. We postulate that persistence represents a more cognitive aspect of motivation compared to the measured intrinsic motivation; especially considering the primary scale of the intrinsic motivation inventory which is interest/enjoyment representing a somewhat affective aspect. Similar to goal-setting, achievements may provide persistence, direction, and effort, but in turn, do not necessarily result in perceiving the task as more interesting and enjoyable. These remarks may be an explanation for why different measures of motivation led to different results.

Regarding performance, results indicate a clear direction. The mean performance of participants with achievements implemented into the task was higher compared to the control groups in all experiments which assessed performance. These results are also in line to similar studies, which have shown other game design elements positively impact performance in an experimental setting; although other elements rather than achievements were implemented (e.g. Attali & Arieli-Attali, 2015; Landers & Landers, 2015; Mekler et al., 2015, 2013). Our results support the proposition that game design elements—specifically achievements—are effective in enhancing performance. We also observed a within-subject improvement in the achievements group over time between a one-week intermittence, whereas the compared groups were lacking this performance improvement. This observation suggests that fully understanding and adapting to an achievement system may take some time. Hence, the full potential of achievements might need some time to develop.

Concluding research question one, achievements are effective in enhancing performance. This enhancement may yet increase over time. Regarding motivation, only persistence showed an increase due to achievements, not in the matter of self-rated interest and enjoyment.

RQ2. How do achievements need to be designed regarding quantity and difficulty to be effective?

Experiment 1 addresses the issue of achievement quantity and its impact on motivation. Although our Hypothesis (H2.1) was not fully supported, a low quantity of achievements led to extended task persistence compared to no achievements, whereas a high quantity of achievements did not. These results suggest preferring a low quantity over a high quantity when deploying achievements. Two explanations for this finding may apply. First, prospect theory (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981) implies accumulating achievements (as a reward) may decrease the individual value of each additional achievement. A decreased value leads to a reduced motivation to unlock further achievements. On the other hand, the value of a few rewards should be higher when individuals have to work longer and harder to achieve them, hence investing more time in reward obtainment. Second, assuming that achievements represent sub-goals with a goal-setting function (Antin & Churchill, 2011; Sailer et al., 2013), the overuse of goals (Ordonez, Schweitzer, Galinsky, & Bazerman, 2009), respectively achievements, may have caused detrimental effects in the high quantity group which resulted in a lacking persistence difference between the high quantity achievements group and the control group. Individuals may have focused more on the easy achievements and felt satisfied with their performance early due to the early high amount of obtained achievements compared to individuals within the low quantity group. This may explain why only the low quantity group showed a higher persistence in the first task than the other groups.

We also found in Experiment 1 that a low quantity of achievements leads to a higher perceived achievement difficulty than a high quantity of achievements. This finding supports our argumentation that a high amount of achievements, based on a hierarchy structure, will lead to the perception that they are easy to achieve. This perception may be due to the early accumulation of low-hierarchy achievements. Hence, a low quantity of achievements should be preferred in achievement-systems over a high quantity, because it leads to a higher perceived difficulty. The benefits of the latter showed in Experiment 2, where results supported our hypothesis that difficult achievements lead to higher performance than easy achievements. They also go in line with our assumption that achievements provide a goal-setting function by formulating clear goals in conjunction with direct behavioral instructions. Goal-setting research has shown over decades a positive relationship between goal difficulty and performance (e.g., Locke & Latham, 1990; Mento et al., 1987, 1980; Yukl & Latham, 1978). Our results indicate that this relationship also may stand true for achievements.

Additionally, we found that participants directly adapted their performance after achievements were introduced. Whereas performance means were identical in the first task (no achievements) for all groups, participants in both achievement groups improved in their performance. The mean of the easy achievement group (15 annotations per picture) aligns to the highest given achievement of also 15 annotations per picture. This supports our assumption that participants align their performance to the highest achievement and do not try to draw upon their full performance potential. In particular, taking into account that the difficult achievement group further outperformed the easy achievement group significantly by three annotations per picture on average in the second task. However, the difficult achievements group did not adjust their performance to the difficult achievement, but rather to the easy achievement in their respective group. We argue they reached their natural performance limit and that the higher achievements in the difficult group were too difficult to achieve. This assumption is supported by the fact that participants indicated they find the achievements very difficult to attain with a score of 4.21 on a five-point scale. However, we do not have further empirical evidence to proof this assumption for certain.

On the other hand, we found no effects for motivation. Groups did not differ between each other regarding self-reported intrinsic motivation measured by interest and enjoyment. As already addressed in the discussion for the first research question, we believe that, although participants perform better with achievements—difficult achievements in particular—they do not enjoy the task more or show more interest in the task. It does not matter if the achievements are easy or difficult to obtain.

Concluding research question two, results suggest to deploy achievements with a high difficulty (though realistic) and with a low quantity to enhance performance. However, interest and enjoyment seem to remain unaffected by the design of the achievements. Only persistence extended when achievements were presented in a low quantity.

RQ3. Do achievements share similar mechanics with classical goal-setting?

Experiment 1 and 2 revealed results which indicate that achievements share similar mechanics to classical goal-setting. First, a low quantity of achievements resulted in a higher perceived achievement difficulty than a high quantity of achievements. This difference can be explained by transferring the argumentation from goal-setting that multiple goals may lead to the tendency to concentrate only on one or a few goals, ignoring the other goals (Shah et al., 2002). Following this

argumentation, we assume that participants concentrated majorly on easy achievements leading to the perception of the high quantity achievement system being easier than the achievement system with fewer achievements. Second, difficult achievements led to a higher performance than easy achievements. This result also supports our assumption that achievements provide a goal-setting function. The positive relationship between goal difficulty and performance is well probed (Locke & Latham, 1990, 2005). The fact that we found the same relationship for achievements underlines the argument that achievements provide clear goals with direct behavioral instructions on how to achieve them. This raise in performance due to difficult goals is explained by the argument that individuals adjust their performance to the difficulty of the goal (Locke, 1996). Third, we found indications that this is also to be true with achievements. The performance mean of participants in the easy achievement group of Experiment 2 almost perfectly matched the highest achievement provided. This matching underlines the adjustment assumption. However, we found only partial support because in the difficult group participants adjusted their performance rather to the lowest achievement instead of the highest achievement. Nevertheless, this adjustment led to a better performance than the group with easy achievements. We believe that the most difficult achievements in the difficult condition were set unrealistically high, which may explain why the occurrence of attaining the difficult achievements showed to be low.

We designed Experiment 3 to directly compare an achievement system to classical goal-setting. We intended to make a first step in investigating the question if achievements are comparable to classical goal-setting. In the matter of self-reported motivation, it has to be noted that both, goal-setting as well as achievements, did not affect motivation. This lack of motivational effects is consistent in Experiment 2 and 3. It may be due to the kind of motivational construct we measured. Compared to Experiment 1, where we assessed a behavioral construct—and found evidence for a motivational benefit of achievements—we measured motivation in Experiment 2 and 3 via the Intrinsic Motivation Inventory in a self-report (Ryan, 1982). Ryan and Deci (2000) distinguish two basic types of motivation, *intrinsic motivation*, “which refers to doing something because it is inherently interesting or enjoyable.” (p.55), and *extrinsic motivation*, “which refers to doing something because it leads to a separable outcome.” (p.55). An explanation for lacking motivational effects may be that achievements naturally do not enhance intrinsic motivation, but only extrinsic motivation. Especially, when keeping in mind that most achievements—as well as the formulated goals—were performance oriented. Many authors argue that performance goals have adverse effects in contrast to mastery goals which are facilitative of intrinsic motivation (Deci et al., 1999; Heyman & Dweck, 1992; Nicholls, 1989). Taking together this assumption that performance oriented achievements and goals harm intrinsic motivation, and the fact that we only measured intrinsic motivation in Experiment 2 and 3, forms a plausible explanation of why no motivational effects revealed in Experiment 2 and 3. In contrast, persistence was measured in Experiment 1. Persistence, as a construct, cannot distinguish between intrinsic and extrinsic motivation. The reason, why an individual decides to persist longer with the task, may be either he or she intrinsically enjoy the task itself or is mainly interested in the outcome externally. The fact that an effect of achievements was found only in Experiment 1 and not in Experiment 2 and 3 supports the assumption that achievements mainly affect extrinsic motivation instead of intrinsic motivation.

Addressing performance, we found significant effects in Experiment 3. The achievement group outperformed the control group but not the goal-setting group. This lacking performance difference between goal and achievement group would support our assumption that achievements and goals share similarities. However, the goal-setting group did not outperform the control group whereas the achievement group did. This lacking difference limits the support for our similarity-assumption. Nevertheless, the obtained medium effect size of $d = 0.56$ for the mean difference between the goal-setting group and the control group is in concordance with other goal-setting research where similar effect sizes have been found (Mento et al., 1987). As we already addressed in Section 7.3, one may argue that the lack of statistical significance may be attributed to the low sample size. A larger sample size might have shown clearer effects of goal-setting, especially regarding the argument that the effect of goal-setting is very reliable and goal-setting theory is among the most valid theories of motivation in organizational psychology (Locke & Latham, 2002). We still argue that the goal-setting effect exists; however we did not find it with our study. Additionally, although we believe in similarities between goal-setting and digital achievements, we also believed that the performance improvement of achievements would increase over time. We found evidence for this assumption in our data. The achievement group was the only group which showed a performance increase from Session 1 to Session 2 which were one-week apart. Several argumentations may explain this phenomenon. First, the deployed achievement system contained a rather high amount of achievements including hierarchical sub-goals. This demand for initial information processing and cognitive load may hinder the full beneficial effects of achievements. Only after a certain adaption time, achievements may reach their full potential of motivational benefits. We only measured two points in time within a one-week interval. If a further increase after this time may occur, is possible, but remains unanswered in our design. Second, achievements may constitute more tangible rewards than conventional goals. Classical goal-setting does not necessarily include rewards; it also did not in our experiments. Digital achievements on the other hand, per definition, include rewards (Montola et al., 2009). The minimum reward of an achievement is always—at least—the grant of a visual representation of an achievement (e.g., badge, trophy). We do not want to get into the controversial debate if tangible rewards undermine motivation (Cameron, Banko, & Pierce, 2001; Deci, Koestner, & Ryan, 2001) at this point. We want to emphasize that we believe digital achievements may be more than just goals with rewards. Different from other rewards, such as monetary incentives or verbal praise, achievements may satisfy certain needs. These needs could be, for example, the desire to collect (Reiss, 2004) or the need for closure (Roets & Hiel, 2011; Webster & Kruglanski, 1994). Individuals may be more satisfied with digital achievements because they do not simply fulfill a given goal, they also can start an (achievement) collection, for example. Furthermore, they may satisfy their need for closure by achieving the full amount of achievements, instead of finishing early. Third, achievements may specifically satisfy intrinsic psychological needs based upon the self-determination theory (Ryan & Deci, 2000). One is the need for competence, the experience of effectiveness in interacting with one's environment (Vansteenkiste & Ryan, 2013). The other one, autonomy, refers to being the perceived origin or source of one's own behavior as well as to experience psychological freedom carrying out an action (Deci & Ryan, 2002; Ryan & Deci, 2000). Sailer et al. (2017) found evidence that game design elements, including achievements, positively affected competence and autonomy need satisfaction. We argue that this need satisfaction may also apply to our study. Due to the design of the achievement system, participants received regular positive feedback. Whenever they fulfilled the requirements of an achievement, they received the reward of a corresponding visual badge or trophy together with a commendatory phrase of how well they are doing. This is in direct contrast to the goal-setting group. They received one goal which were identical to the highest achievement. No mid-task commendation was given in this group. Although all groups received constant feedback about the amount of generated ideas, the feel of competence may be highest in the achievements group. The design of the achievement system might also have affected the feel of autonomy. Achievements formed several categories, such as achievements for speed, quantity or meta-achievements (collecting a certain amount of certain achievements). This might evoke a certain feel of autonomy or strategic decision making. Instead of merely generating a high amount of ideas,

participants might reflect about how to specifically approach which achievement. For example, one achievement was only available within the first 5 min of a task, another one was only available within the last minute of a task. This might have led to potential deliberate decisions which fostered the feel of autonomy and psychological freedom. However, these remain speculative consideration, as we did not measure the motives comprised in self-determination theory specifically.

After given some explanations, of why digital achievements outperformed goal-setting in the second session, we now conclude our third research question. We found that a low quantity of achievements as well as achievements with a high difficulty foster performance. Additionally, we found indications that individuals adapt their performance directly along certain achievements. These results are concordant to results and argumentations of conventional goal-setting literature with regards to pre-formulated goals (Locke, 1996; Locke & Latham, 1990, 2005; Shah et al., 2002). We conclude that achievements share mechanics similar to classical goal-setting. However, this does not mean they are equally interchangeable with goals. In a direct comparison, we even found an outperformance of achievements over classical goals over time. We believe that achievements can satisfy certain needs, such as the need for competence, autonomy or cognitive closure, better than classical goal-setting.

8.2. Limitations and prospects

We will now present some limitations of our studies and prospects of future gamification research. First, we intend to address generalization. In our experiments, participants interacted only for a relatively brief period with the game design elements. The time span ranged from a couple of minutes up to one and half an hour. Also, with a Stroop task, image annotation, and idea generation, the deployed tasks were rather context-free without a profound purpose. Furthermore, we focused on a single game design element instead of an entire gamification system. All these aspects may be criticized for lack of transferability to proper *real-life* applications of gamification. Most of the commonly used wide-stretched gamification systems implemented into an app or a platform, users interact with over weeks up to month fulfilling a purposeful task or goal. However, we deliberately opt for this design to tackle some of the main caveats of prior gamification research. Our intention was to unconfound game design elements from each other in order to derive specific conclusions about achievements, as well as to remove the high specific context of other studies. Our experimental approach also allows for more control and a higher internal validity compared to a more field-oriented approach.

Furthermore, although we found evidence for similarities between digital achievements and classical goal-setting, our conclusions are not without limits. We did not investigate the specific constructs which are reckoned to be the underlying working mechanisms behind the relationship between goal-setting and motivation, respectively performance. These might be for example direction or effort (Locke & Latham, 2002). However, this was not the main focus of the respective experiments. We intended to investigate the design of achievements, which showed to hold similarities to the design of conventional goals. We suggest that future research should focus more on studies with a direct comparison between achievements and goal-setting as we did in our third experiment.

Another limitation might be that some constructs were only measured with a single item, such as achievement difficulty or task complexity in Experiment 1. Single-item measures are often criticized because of their assumed low reliability and validity. However, we took these single-measurements out of common questionnaires and omitted the rest intentionally due to lacking compatibility with our context or economical reasons (i.e., tradeoff between importance of the measure and length of the questionnaire). Also, the single-item measures served the purpose of secondary control variables rather than contributing to the major research objectives.

Gamification research still is in its early stages. There are many open questions and aspects which are exciting and important to investigate to paint a clearer picture of the use and explanation regarding the topic of gamification. We now intend to give promising prospects for research ideas, which we derive out of the studies of this work.

First, we propose to investigate other measures for motivation. We reckon it as essential to shed more light especially on the discussion whether gamification fosters extrinsic or intrinsic motivation (Seaborn & Fels, 2015). We believe that single elements rather address extrinsic motivation whereas complex gamification systems foster intrinsic motivation. This differentiation, however, is highly dependent on the design of the game design elements deployed. Unfortunately, gamification research has highly neglected the design of gamification. We investigated some design aspects of achievements in our studies; however, there are many other aspects of which we do not know how they influence the gamification experience. With regard to achievements, further aspects may be of interest, such as timing, visual presentation, phrasing, or fulfillment logic. As we discussed in earlier sections, achievement design is much more multilayered than achievements present versus absent. Furthermore, we propose to extend the single element approach to other elements. Although external validity might be constricted, it is necessary to obtain clear and unconfounded knowledge about how single game design elements work and how they should be designed to foster motivation and performance effectively. Another large somewhat uncharted field addresses user characteristics. Although studies were conducted with various samples, direct comparisons based upon user characteristics have been rather neglected. We believe that individuals might experience the same gamification system profoundly different, dependent on personality traits or individual attitudes. These could be, for example, achievement motivation, need for cognitive closure, task importance or conscientiousness. Last, more research should investigate the task itself and inherent task related characteristics, which might influence the outcome of gamification. For example, different learning ability levels, required to handle a task (i.e., passive, active, constructive, and interactive; Chi, 2009), or variations of performance oriented versus mastery oriented tasks (Dweck, 1986; Nicholls, 1984) might change the effectiveness of certain game design elements.

9. Conclusion

In the three experiments of this paper, we focused on digital achievements as a cornerstone of gamification systems. We investigated achievements in a controlled experimental environment unconfounded from other game design elements. Our three research questions intended to answer, if achievements are effective in enhancing motivation and performance, how they need to be designed to be effective, and, if achievements resemble classical goal-setting. Results strongly suggest that achievements have the potential to enhance performance; this enhancement strengthens over time. Results of motivation were more controversial. Whereas persistence was higher with achievements, self-reported intrinsic motivation was consistent throughout all groups. Results also suggest to deploy achievements with low quantity and high difficulty, which showed to raise performance. Achievements also share mechanics similar to classical goal-setting; they even outperformed goal-setting over time. We argue that achievements satisfy specific needs better than conventional goal-setting, such as the need for competence, autonomy or cognitive closure.

We see gamification as a potentially powerful tool in evoking motivational and performance-related benefits. Further research needs to continue to investigate the single aspects of promising game design elements and systems. We believe it is not a question of *to do or not to do?*, but rather of *how to do it right?*

Author note

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declarations of interest: none. Correspondence concerning this article should be addressed to Christopher Groening, University of Muenster, Institute of Psychology - Work Psychology, Fliednerstr. 21, 48,149 Muenster, Germany. E-mail address: christopher.groening@wwu.de.

Acknowledgements

We thank Paula Blumenschein, Fiona Kazarovytska, Kristina L. van Vorst, Karina Kühn and Dominik Sondern for their help in study design and data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2019.02.026.

References

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). Steering user behavior with badges. *Proceedings of the 22nd international Conference on world wide web* (pp. 95–106). ACM Press. https://doi.org/10.1145/2488388.2488398.

Antin, J., & Churchill, E. F. (2011). Badges in social media: A social psychological perspective. *Chi 2011 gamification workshop proceedings (vancouver, bc, Canada, 2011)*.

Aparicio, A. F., Vela, F. L. G., Sanchez, J. L. G., & Montes, J. L. I. (2012). Analysis and application of gamification. *Proceedings of the 13th international conference on interacción persona-ordenador*New York, NY, USA: ACM. https://doi.org/10.1145/2379636.2379653 17:1-17:2.

Arai, S., Sakamoto, K., Washizaki, H., & Fukazawa, Y. (2014). A gamified tool for motivating developers to remove warnings of bug pattern tools. In *2014 6th international workshop on empirical software engineering in practice. IEEE*. https://doi.org/10.1109/iwesep.2014.17.

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education, 83*, 57–63. https://doi.org/10.1016/j.compedu.2014.12.012.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*, 307–337.

Barata, G., Gama, S., Jorge, J., & Goncalves, D. (2013). Engaging engineering students with gamification. *2013 5th international conference on games and virtual worlds for serious applications*IEEEhttps://doi.org/10.1109/vs-games.2013.6624228.

Bassett, G. A. (1979). A study of the effects of task goal and schedule choice on work performance. *Organizational Behavior & Human Performance, 24*(2), 202–227. https://doi.org/10.1016/0030-5073(79)90026-6.

Bryan, J. F., & Locke, E. A. (1967). Goal setting as a means of increasing motivation. *Journal of Applied Psychology, 51*(3), 274–277. https://doi.org/10.1037/h0024566.

Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation: The myth continues. *The Behavior Analyst, 24*(1), 1–44. https://doi.org/10.1007/bf03392017.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science, 1*(1), 73–105.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105–115. https://doi.org/10.1037/h0030644.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62*(1), 119–142. https://doi.org/10.1111/j.1467-6494.1994.tb00797.x.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668. https://doi.org/10.1037/0033-2909.125.6.627.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1–27. https://doi.org/10.3102/00346543071001001.

Deci, E. L., & Ryan, R. M. (2002). Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research*, 3–33.

Denny, P. (2013). The effect of virtual achievements on student engagement. *Proceedings of the SIGCHI conference on human factors in computing systems - CHI 13*ACM Presshttps://doi.org/10.1145/2470654.2470763.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011a). From game design elements to gamefulness. *Proceedings of the 15th international academic MindTrek conference on envisioning future media environments - MindTrek: Vol. 11*ACM Presshttps://doi.org/10.1145/2181037.2181040.

Deterding, S., Khaled, R., Nacke, L. E., & Dixon, D. (2011b). Gamification: Toward a definition. *Chi 2011 gamification workshop proceedings*.

Dickey, M. D. (2006). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research & Development, 55*(3), 253–273. https://doi.org/10.1007/s11423-006-9004-7.

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of Personality and Social Psychology, 61*(3), 392–403. https://doi.org/10.1037/0022-3514.61.3.392.

Domínguez, A., de Navarrete, J. S., de Marcos, L., Fernández-Sanz, L., Pages, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education, 63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020.

Donato, P., & Michael, W. (2013). Link. 2013"the gamification of marketing research". *Marketing News*, 38–42.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 66*(5), 968.

Entertainment Software Association (2017). *Essential facts about the computer and video game industry (resreport)*. Entertainment Software Association. Retrieved from http://essentialfacts.theesa.com/mobile/.

Farzan, R., DiMicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., & Dugan, C. (2008). When the experiment is over: Deploying an incentive system to all the users. *Symposium on persuasive technology*.

Feather, N. T. (1961). The relationship of persistence at a task to expectation of success and achievement related motives. *Journal of Abnormal and Social Psychology, 63*(3), 552–561. https://doi.org/10.1037/h0045671.

Fernandes, J., Duarte, D., Ribeiro, C., Farinha, C., Pereira, J. M., & da Silva, M. M. (2012). iThink: A game-based approach towards improving collaboration and participation in requirement elicitation. *Procedia Computer Science, 15*, 66–77. https://doi.org/10.1016/j.procs.2012.10.059.

Fitz-Walter, Z., Tjondronegoro, D., & Wyeth, P. (2011). Orientation passport: Using gamification to engage university students. *Proceedings of the 23rd australian computer-human interaction conference on - OzCHI: Vol. 11*ACM Presshttps://doi.org/10.1145/2071536.2071554.

Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal, 35*(2), 350–369. https://doi.org/10.5465/256377.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467. https://doi.org/10.1177/1046878102238607.

Girdler, S. S., Turner, J. R., Sherwood, A., & Light, K. C. (1990). Gender differences in blood pressure control during a variety of behavioral stressors. *Psychosomatic Medicine, 52*(5), 571–591. https://doi.org/10.1097/00006842-199009000-00009.

Goh, D. H., & Lee, C. S. (2011). Perceptions, quality and motivational needs in image tagging human computation games. *Journal of Information Science, 37*(5), 515–531. https://doi.org/10.1177/0165551511417786.

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *American Psychologist, 69*(1), 66–78. https://doi.org/10.1037/a0034857.

Guin, T. D.-L., Baker, R., Mechling, J., & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research, 54*(5), 613–633. https://doi.org/10.2501/ijmr-54-5-613-633.

Hakulinen, L., Auvinen, T., & Korhonen, A. (2013). Empirical study on the effect of achievement badges in trakla2 online learning environment. *Learning and teaching in computing and engineering (latice)* (pp. 47–54). 2013.

Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The effect of achievement badges on students' behavior: An empirical study in a university-level computer science course. *International Journal of Emerging Technologies in Learning (iJET), 10*(1), 18–29.

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*(4), 236–245. https://doi.org/10.1016/j.elerap.2013.01.004.

Hamari, J., & Eranti, V. (2011). Framework for designing and evaluating game achievements. In: *Proc. DiGRA 2011: Think Design Play, 115*, 122–134.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - a literature review of empirical studies on gamification. *2014 47th Hawaii international conference on system sciences*IEEEhttps://doi.org/10.1109/hicss.2014.377.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019.

Hense, J., Klevers, M., Sailer, M., Horenburg, T., Mandl, H., & Günthner, W. (2014). Using gamification to enhance staff motivation in logistics. *Frontiers in gaming simulation* (pp. 206–213). Springer International Publishing. https://doi.org/10.1007/978-3-319-04954-0_24.

Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion, 16*(3), 231–247. https://doi.org/10.1007/bf00991653.

Hollenbeck, J. R., & Klein, H. J. (1987). Goal commitment and the goal-setting process: Problems, prospects, and proposals for future research. *Journal of Applied Psychology, 72*(2), 212–220. https://doi.org/10.1037/0021-9010.72.2.212.

Hollenbeck, J. R., Klein, H. J., O'Leary, A. M., & Wright, P. M. (1989). Investigation of the construct validity of a self-report measure of goal commitment. *Journal of Applied Psychology, 74*(6), 951–956. https://doi.org/10.1037/0021-9010.74.6.951.

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the stroop color and word test with children. *Archives of Clinical Neuropsychology, 19*(6), 725–743. https://doi.org/10.1016/j.acn.2003.09.003.

Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal

setting in group collaboration environments. *Management Science, 56*(4), 724–742. https://doi.org/10.1287/mnsc.1090.1129.
Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350. https://doi.org/10.1037/0003-066x.39.4.341.
Klein, H. J., Wesson, M. J., Hollenbeck, J. R., Wright, P. M., & DeShon, R. P. (2001). The assessment of goal commitment: A measurement model meta-analysis. *Organizational Behavior and Human Decision Processes, 85*(1), 32–55. https://doi.org/10.1006/obhd.2000.2931.
Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179–188. https://doi.org/10.1016/j.chb.2014.03.007.
Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191–210.
Landers, R. N. (2014). Developing a theory of gamified learning. *Simulation & Gaming, 45*(6), 752–768. https://doi.org/10.1177/1046878114563660.
Landers, R. N., & Landers, A. K. (2015). An empirical test of the theory of gamified learning. *Simulation & Gaming, 45*(6), 769–785. https://doi.org/10.1177/1046878114563662.
Lang, J. W., & Fries, S. (2006). A revised 10-item version of the achievement motives scale. *European Journal of Psychological Assessment, 22*(3), 216–224. https://doi.org/10.1027/1015-5759.22.3.216.
LaPorte, R. E., & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology, 68*(3), 260–264. https://doi.org/10.1037/0022-0663.68.3.260.
Latham, G. P. (2015). Motivate employee performance through goal setting. *Handbook of principles of organizational behavior* (pp. 161–178). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119206422.ch9.
Latham, G. P., & Locke, E. A. (1979). Goal setting-a motivational technique that works. *Organizational Dynamics, 8*(2), 68–80. https://doi.org/10.1016/0090-26167990032-9.
Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*(2), 212–247. https://doi.org/10.1016/0749-5978(91)90021-k.
Latham, G. P., Mitchell, T. R., & Dossett, D. L. (1978). Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. *Journal of Applied Psychology, 63*(2), 163.
Latham, G. P., & Saari, L. M. (1979). Importance of supportive relationships in goal setting. *Journal of Applied Psychology, 64*(2), 151–156. https://doi.org/10.1037/0021-9010.64.2.151.
Lee, C., & Bobko, P. (1992). Exploring the meaning and usefulness of measures of subjective goal difficulty 1. *Journal of Applied Social Psychology, 22*(18), 1417–1428. https://doi.org/10.1111/j.1559-1816.1992.tb00957.x.
Liu, Y., Alexandrova, T., & Nakajima, T. (2011). Gamifying intelligent environments. *Proceedings of the 2011 international ACM workshop on ubiquitous meta user interfaces - ubi-MUI: Vol. 11*ACM Presshttps://doi.org/10.1145/2072652.2072655.
Locke, E. A. (1966). The relationship of intentions to level of performance. *Journal of Applied Psychology, 50*(1), 60–66. https://doi.org/10.1037/h0022937.
Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior & Human Performance, 3*(2), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4.
Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology, 5*(2), 117–124. https://doi.org/10.1016/s0962-1849(96)80005-9.
Locke, E. A., & Bryan, J. F. (1969). The directing function of goals in task performance. *Organizational Behavior & Human Performance, 4*(1), 35–42. https://doi.org/10.1016/0030-5073(69)90030-0.
Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology, 69*(2), 241–251. https://doi.org/10.1037/0021-9010.69.2.241.
Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. *Choice Reviews Online, 28*(01), https://doi.org/10.5860/choice.28-0608 28-0608-28-0608.
Locke, E. A., & Latham, G. (1994). *Goal-setting theory. Organizational behavior 1: Essential theories of motivation and leadership.* 159–183.
Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. https://doi.org/10.1037/0003-066x.57.9.705.
Locke, E. A., & Latham, G. (2005). Goal-setting theory. In J. B. Miner (Ed.). *Organizational behavior 1: Essential theories of motivation and leadership* (pp. 159–183). Routledge. https://doi.org/10.4324/9781315702018.
Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current Directions in Psychological Science, 15*(5), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x.
Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin, 90*(1), 125–152. https://doi.org/10.1037/0033-2909.90.1.125.
Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International journal of management, business, and administration, 15*(1), 1–6.
MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203. https://doi.org/10.1037/00332909.109.2.163.
de Marcos, L., Domínguez, A., de Navarrete, J. S., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education, 75*, 82–91. https://doi.org/10.1016/j.compedu.2014.01.012.
Markets and Markets (2016). *Gamification market by solution (consumer driven and enterprise driven), applications (sales and marketing), deployment type (on-premises and cloud), user type (large enterprise, smbs), industry and region - global forecast to 2020 (resreport) . Markets and Markets.* Retrieved from https://www.researchandmarkets.com/reports/3618296/gamification-market-by-solution-consumer-driven.
Martin, J. J., & Gill, D. L. (1995). The relationships of competitive orientations and self-efficacy to goal importance, thoughts, and performance in high school distance runners. *Journal of Applied Sport Psychology, 7*(1), 50–62. https://doi.org/10.1080/10413209508406300.
Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance, 10*(4), 303–330. https://doi.org/10.1207/s15327043hup1004_1.
McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise & Sport, 60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413.
McDaniel, R., Lindgren, R., & Friskics, J. (2012). Using badges for shaping interactions in online learning environments. *2012 IEEE international professional communication conference* (pp. 1–4). IEEE. https://doi.org/10.1109/ipcc.2012.6408619.
Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Disassembling gamification. *CHI 13 extended abstracts on human factors in computing systems on - CHI EA '13*ACM Presshttps://doi.org/10.1145/2468356.2468559.
Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2015). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior, 71*, 525–534. https://doi.org/10.1016/j.chb.2015.08.048.
Mento, A. J., Cartledge, N. D., & Locke, E. A. (1980). Maryland vs Michigan vs Minnesota: Another look at the relationship of expectancy and goal difficulty to task performance. *Organizational Behavior & Human Performance, 25*(3), 419–440. https://doi.org/10.1016/0030-5073(80)90038-0.
Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966-1984. *Organizational Behavior and Human Decision Processes, 39*(1), 52–83. https://doi.org/10.1016/0749-5978(87)90045-8.
Mollick, E. R., & Rothbard, N. (2013). Mandatory fun: Gamification and the impact of games at work. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2277103.
Montola, M., Nummenmaa, T., Lucero, A., Boberg, M., & Korhonen, H. (2009). Applying game achievement systems to enhance user experience in a photo sharing service. *Proceedings of the 13th international MindTrek conference: Everyday life in the ubiquitous era on - MindTrek '09*ACM Presshttps://doi.org/10.1145/1621841.1621859.
Muntean, C. I. (2011). Raising engagement in e-learning through gamification. In: *Proc. 6th international conference on virtual learning icvl* (pp. 323–329). .
Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience , task choice, and performance. *Psychological Review, 91*(3), 328.
Nicholls, J. G. (1989). The competitive ethos and democratic education. *Choice Reviews Online, 27*(02), https://doi.org/10.5860/choice.27-1049 27-1049-27-1049.
Nicholson, S. (2012). A user-centered theoretical framework for meaningful gamification. *Games+ Learning+ Society, 8*.
Ordonez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals gone wild: The systematic side effects of overprescribing goal setting. *Academy of Management Perspectives, 23*(1), 6–16. https://doi.org/10.5465/amp.2009.37007999.
Paulus, P. B., & Dzindolet, M. T. (1993). Social influence processes in group brainstorming. *Journal of Personality and Social Psychology, 64*(4), 575–586. https://doi.org/10.1037/0022-3514.64.4.575.
Peng, W., Lin, J.-H., Pfeiffer, K. A., & Winn, B. (2012). Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psychology, 15*(2), 175–196. https://doi.org/10.1080/15213269.2012.673850.
Popović, M. D., Kostić, M. D., Rodić, S. Z., & Konstantinović, L. M. (2014). Feedback-mediated upper extremities exercise: Increasing patient motivation in poststroke rehabilitation. *BioMed Research International,* 1–11. https://doi.org/10.1155/2014/520374 2014.
Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154–166. https://doi.org/10.1037/a0019440.
Reinecke, L. (2009). Games at work: The recreational use of computer games during working hours. *CyberPsychology and Behavior, 12*(4), 461–465. https://doi.org/10.1089/cpb.2009.0010.
Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology, 8*(3), 179–193. https://doi.org/10.1037/1089-2680.8.3.179.
Roets, A., & Hiel, A. V. (2011). Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and Individual Differences, 50*(1), 90–94. https://doi.org/10.1016/j.paid.2010.09.004.
Rothkopf, E. Z., & Billington, M. J. (1979). Goal-guided learning from text: Inferring a descriptive processing model from inspection times and eye movements. *Journal of Educational Psychology, 71*(3), 310–327. https://doi.org/10.1037/0022-0663.71.3.310.
Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450–461. https://doi.org/10.1037/0022-3514.43.3.450.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066x.55.1.68.
Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *IxD&A, 19*, 28–37.
Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006.

Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*(6), 1261–1280. https://doi.org/10.1037/0022-3514.83.6.1261.

Smith, C. P. (1964). Relationships between achievement-related motives and intelligence, performance level, and persistence. *Journal of Abnormal and Social Psychology, 68*(5), 523–532. https://doi.org/10.1037/h0045465.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651.

Su, C.-H., & Cheng, C.-H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning, 31*(3), 268–286. https://doi.org/10.1111/jcal.12088.

Turner, J., Hewitt, J. K., Morgan, R. K., Sims, J., Carroll, D., & Kelly, K. A. (1986). Graded mental arithmetic as an active psychological challenge. *International Journal of Psychophysiology, 3*(4), 307–309. https://doi.org/10.1016/0167-8760(86)90039-5.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1126/science.7455683.

Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration, 23*(3), 263–280. https://doi.org/10.1037/a0032359.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049–1062. https://doi.org/10.1037/0022-3514.67.6.1049.

Werbach, K. (2014). (re) defining gamification: A process approach. *International conference on persuasive technology* (pp. 266–272). .

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business.* Wharton Digital Press.

Witt, M., Scheiner, C., & Robra-Bissantz, S. (2011). *Gamification of online idea competitions: Insights from an explorative case.* Informatik schafft Communities192.

Wood, R. E. (1990). Goal setting and strategy effects on complex tasks. *Research in Organizational Behavior, 12*, 73–110.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology and Behavior, 9*(6), 772–775. https://doi.org/10.1089/cpb.2006.9.772.

Yukl, G. A., & Latham, G. P. (1978). Interrelationships among employee participation, individual differences, goal difficulty, goal acceptance, goal instrumentality, and performance. *Personnel Psychology, 31*(2), 305–323. https://doi.org/10.1111/j.1744-6570.1978.tb00449.x.

Zicherman, G., & Cunningham, C. (2011). *Gamification by design.* CA: O'Reilly Media.