Journal of Aging and Physical Activity, 2024, 32, 761-770
https://doi.org/10.1123/japa.2022-0249

Human Kinetics
ORIGINAL RESEARCH

Dosed Failure Increases Older Adult's Motivation for an Exergame

Nick Kluft, Jeroen B.J. Smeets, and Katinka van der Kooij
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

We investigated whether dosed failure motivates older adults to perform more repetitions in an exergame that involves hitting targets with stepping movements. The effect of dosed failure was studied in a within-participants design in which all participants performed this exergame in both a Standard condition, in which one never fails, and in a Dosed Failure condition, in which we introduced about 30% failures. The order of conditions (Standard First or Dosed Failure first) was chosen randomly for each participant. Results showed that participants performed more repetitions in the Dosed Failure condition compared with the Standard condition, while play duration and subjective motivation at the moment of quitting did not differ. This shows that dosed failure motivated older adults to put a greater amount of effort to perform the exercise without affecting play duration or subjective motivation.

Keywords: incentives, achievement, serious games, flow

Key Points

- Older adults put greater effort into an exergame when they sometimes fail compared with when they always receive success feedback.
- Introducing some failures did not affect self-reported motivation.

Physical exercise is important for older adults, who are advised to engage in at least 150 min of moderate physical activity per week and in addition biweekly muscle- and bone-strengthening exercises (World Health Organization, 2020). Unfortunately, not everyone is motivated by physical exercise. Computer games provide an opportunity to design motivating physical exercises ("exergames"). For instance, a challenge can be personalized to the individual's performance.

Success is motivating (Burgers et al., 2015; Cook & Artino, 2016; Ryan & Deci, 2002; Wulf & Lewthwaite, 2016). However, this does not mean that making a task very easy is the right way to motivate someone. In the present study, we asked whether failing once in a while increases older adults' motivation for an exergame. Two prominent theories of motivation, Flow Theory (Csikszentmihalyi, 1990) and Achievement Motivation Theory (Atkinson, 1957), provide a framework for understanding why such dosed failure might be motivating.

In the Flow Theory, humans enter a "Flow" state of optimal motivation and focus when challenge and skill are in balance (Csikszentmihalyi, 1990). Too much challenge results in anxiety, whereas too little challenge results in boredom. Therefore, the relationship between the motivation and the challenge follows an inverted-U. There is ample evidence in a variety of game types that illustrates the inverted-U relationship between the motivation and the challenge (for a review, see Fong et al., 2015). For instance, games of chess against a superior opponent are rated as more motivating than a games of chess against an inferior opponent and close wins are more enjoyable than blowouts (Abuhamdeh & Csikszentmihalyi, 2012). Although Flow Theory (Csikszentmihalyi, 1990) did not express challenge in terms of successes and failures, we propose that never failing corresponds to a lack of challenge, whereas always failing corresponds to an excess of challenge. Hence, the optimal challenge must involve some failures.

Besides signaling the optimal challenge, dosed failure might be motivating because it enhances the value of success. In Achievement Motivation Theory (Atkinson, 1957), the motivation is a product of the probability of success and the value of success, which decreases with the probability of success. The optimal motivation occurs at the intersection of the probability and the value function (Atkinson, 1957). Originally, a success probability of .5 was proposed as optimally motivating (Atkinson, 1957). However, the optimal success probability depends on the shape of the value function and might be higher than .5. In the context of sports coaching (Burton & Raedeke, 2008) and education (Rosenshine & Stevens, 1986), an 80% success rule has been proposed as an index of the optimal challenge.

Based on the predictions by Achievement Motivation Theory (Atkinson, 1957) and the 80% success rule (Burton & Raedeke, 2008; Rosenshine & Stevens, 1986), older adults might be motivated for physical exercise by designing the challenge of an exergame such that they succeed on 50%–90% of attempts, but fail on the other attempts. Two studies assessed the influence of the challenge on motivation in older adults. These studies used a timed sequential key-pressing task in which the challenge was manipulated by the order of sequences of timing goals: blocked (AAABBBCCC; low challenge), serial (ABCABCABC; moderate challenge), and random (CBABACACB; high challenge). The results showed that with older adults' self-reported motivation was higher in a "learner adapted" condition that changed to a more difficult practice mode when they failed on less than 33% of the

Kluft is now at Dijklander Ziekenhuis, Hoorn, The Netherlands.
Kluft https://orcid.org/0000-0001-6504-8724
Smeets https://orcid.org/0000-0002-3794-0579
van der Kooij (k.vander.kooij@vu.nl) is corresponding author, https://orcid.org/0000-0002-4190-7827

trials, compared with when they always practiced in the low-challenge blocked practice mode or in the high-challenge random mode (Beik & Fazeli, 2021). A later study confirmed these results using electroencephalogram activity as a measure of motivation (Beik et al., 2022). However, when a task requires more physical effort than key pressing, more success might be needed to compensate for the physical effort. Moreover, the finding that the learner-adapted condition was most motivating (Beik et al., 2022) might also be explained by the variety in the timing goals, which were presented on the computer screen, rather than by the resultant success frequency.

Thus, on a theoretical level, we propose that dosed failure might be motivating because it signals the balance between challenge and skill that is key to Flow Theory (Csikszentmihalyi, 1990) and because the value of success might be optimal for a success rate containing some failures (Atkinson, 1957). Experimentally, it has been shown that moderate challenge is motivating (see Fong et al., 2015, for a review), and that motivation is highest in a condition in which the challenge increased when the failure rate exceeded 33% (Beik & Fazeli, 2021). It is unclear, however, whether this motivating effect of dosed failure also holds for tasks involving greater physical effort than button presses.

In this study, we test the hypothesis that instances of failure motivate older adults to engage in an exergame which requires physical effort. For this, we use an exergame developed by SilverFit B.V. in which participants are asked to hit moles displayed on a screen by making stepping movements that are registered by a 3D camera. We manipulate the challenge by pacing the task. The (original) standard version of this exergame involves no experiences of failure, as the next mole only appeared when the previous one had been hit (Standard condition). We created a version of the exergame, in which we introduced a dose of failures by limiting the time in which a mole could be hit, so that participants on average fail to hit the moles on about 30% of the repetitions (Dosed Failure condition). In line with Brehm and Self (1989), we measured motivation objectively from the effort a person voluntarily puts into a task (Brehm & Self, 1989), by counting the number of repetitions the participant played voluntarily. Because individuals differ in the extent to which they are motivated by success (Atkinson, 1957; Ryan & Deci, 2002), we assessed the effect of dosed failure in a within-participant design in which all participants played the exergame in both the Standard condition and in the Dosed Failure condition. We predicted that participants would play more repetitions in the Dosed Failure condition compared with the Standard condition.

Methods

Preregistration

The hypotheses and methods, including the sample plan, of this study were preregistered. Our preregistration can be found in our OpenScienceFramework project page (https://osf.io/bdkzv). We indicated in the "Deviations from Preregistration" section the changes made after this registration.

Participants

We recruited 39 participants from activity classes for older adults by spreading flyers and by word of mouth; for reasons given below, we analyzed the data of only 38 participants (27 females and 11 males, see Table 1 for sample descriptives). Inclusion criteria for participants included 65 years old or older (age 72.1 ± 5.1; mean ± *SD*), reported that they had normal or corrected to normal vision, had a minimental state examination (Folstein et al., 1975) of 20 or higher, were adequately vaccinated against the coronavirus, reported that they could independently walk 20 m consecutively without experiencing shortness of breath or feeling a sharp pain in the chest, reported that they were not diagnosed with either epilepsy or Parkinson's disease, and were not recovering from a cerebrovascular attack or surgery in the lower extremities. Prior to any experimentation, all procedures were explained and participants provided their consent by signing an informed consent form. The local ethics committee Vaste Commissie Wetenschappelijke Ethiek approved the design and all the procedures within the study (VCWE-S-21-001155). We excluded the data of the first participant from the analysis because we adapted our planned procedure to determine the initial pacing of the exergame after this participant (see "Deviations from Preregistration" section). To describe participants' frailty, we measured handgrip strength following the procedure of Reijnierse et al. (2017) and the Dutch version of the Falls Efficacy Scale-International (FES-i) (Kempen et al., 2007; Yardley et al., 2005).

Table 1 **Sample Descriptives**

Variable	Unit	Value	Statistic
Reported age	Years	75 ± 5	*M* ± *SD*
$\text{Sex}_{\text{females}}$	Count	27 (73%)	*N* (%)
MMSE	Points	28 ± 2	*M* ± *SD*
FES-i	Points	21 ± 4	*M* ± *SD*
$\text{HGS}_{\text{females}}$	N	265 ± 39	*M* ± *SD*
$\text{HGS}_{\text{males}}$	N	421 ± 83	*M* ± *SD*

Note. In case a subscript is given, the measure is calculated for a specific subset (i.e., males or females). Mean value and *SD* (*M* ± *SD*) or number and percentage of sample (*N* [%]) are depicted. MMSE = minimental state examination; FES-i = Falls Efficacy Scale-International; HGS = handgrip strength.

Design

To optimize the statistical power, the effect of dosed failure on the motivation was measured in a within-participants design in which we compared the number of repetitions participants voluntarily made between a Standard and a Dosed Failure condition. The order of conditions was chosen randomly for each participant. Twenty-one participants were assigned to a Standard-first group which performed the Standard condition first (reported age = 71.3 ± 4.6 years; 17 females); 17 participants performed were assigned to a Dosed Failure-first group which performed the Dosed Failure condition first (mean age = 73.5 ± 5.7 years; 11 females).

Exergame

We used an existing exergame, called *molegame*, on a SilverFit 3D system (SilverFit). The exergame was played using a demo set-up including a computer, a projector, and a portable screen for visual display, and a 3D camera (Microsoft Kinect) that detected and recorded the 3D location of the participant's feet. Participants stood 3 m in front of the screen and had to tap with their feet at certain positions on the floor to chase moles on the screen with their virtual foot that was also presented on the screen (see Figure 1). Users of this exergame generally play for about 3 min, as indicated by the developer SilverFit B.V.

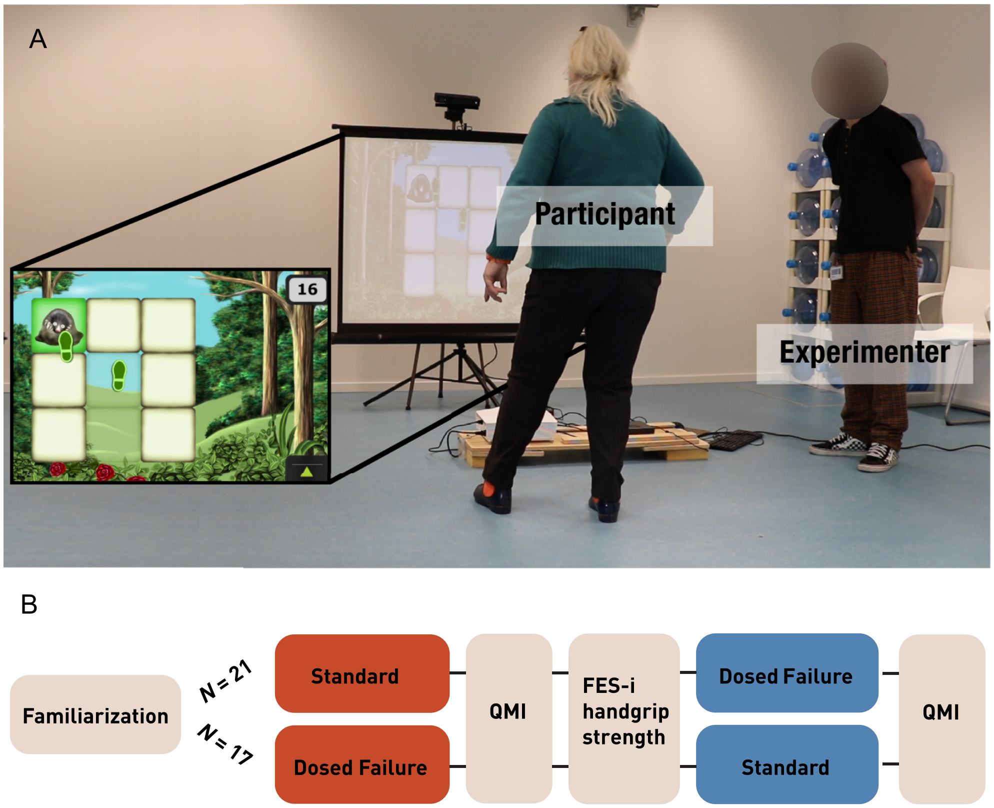

Figure 1 — (A) Experimental set-up. The playing field consisted of seven tiles. The size of the field was adjusted to the participant's leg length. The participant's task was to hit the target (i.e., mole) as soon as possible. The screen provided feedback of the participant's feet and a counter on the top-right corner monitored the total number of targets hit. The experimenter stood near the participant to assist during a potential balance loss to avoid falls. (B) Schematic diagram of the experimental procedures. All the participants received a familiarization trial after which they were divided into two streams, such that 21 participants played the Standard condition first and 21 participants that played the Dosed Failure condition first. After each condition, self-reported motivation was assessed using the QMI. Before the second condition, we assessed the handgrip strength and falls efficacy. FES-i = Falls Efficacy Scale-International; QMI = Quick Motivation Index.

The on-screen playing field consisted of seven square tiles of equal size, on which the mole could appear (see Figure 1). In the original exergame, a 3×3 tile grid is used. For this experiment, the center and center back tiles were removed. We removed these tiles because the registration of the feet by the 3D camera was suboptimal at these locations. The area on the floor that corresponded to the playing field was adjusted to the leg length of the participant, such that the legs made an angle of 30° in an anteroposterior direction and 38° in the mediolateral direction when hitting the mole with one foot, while the other foot remained at the center of the playing field. This ensured that the spatial challenge was similar across individuals: one step from the center of the playing field sufficed to hit each mole. The size of the virtual tiles on the playing field was about 0.4×0.3 m^2, depending on leg size. A counter on the top-right corner of the screen showed the total amount of moles hit within the condition (the score). Immediately after a mole was hit, a new mole randomly appeared on one of the seven tiles.

The participant played this exergame in two conditions: (a) Standard condition, in which there is no time limit to hitting the mole as it remains on the playing field until hit; and (b) Dosed Failure condition, in which the mole disappears after a given time (t_{mole}).

In the Standard condition, the mole remained visible until hit, so that the participant was always successful in hitting the mole. When the participant hit the mole, an exit animation of the mole was shown, an encouraging sound was played, and the underlying tile turned green.

In the Dosed Failure condition, we paced the exergame, such that participants would fail hitting the mole in about 30% of the repetitions, by limiting the time the participant has to hit the mole. The participant could get two kinds of feedback from the game. When they hit the mole in time, the feedback was the same as in the Standard condition. When the mole was not hit in time, it disappeared showing the exit animation without the tile turning green or playing the encouraging sound. We ensured that participants failed in about 30% of the repetitions by carefully selecting the time the participant had to hit a mole (t_{mole}). For each participant, we determined an individual initial t_{mole} during a familiarization period (see "Procedure" section). We adjusted the t_{mole} value during the experiment based on the success frequency. When a participant hit eight or more moles out of the previous nine repetitions (success frequency ≥0.9), the t_{mole} decreased by 3% and when the participant failed on four or more targets (success frequency ≤0.5), the t_{mole} increased by 3%. We selected these percentages and the number of repetitions used to determine the success frequency (i.e., nine targets) based on pilot studies.

Procedure

After the participant signed the informed consent, we administered the minimental state examination, and measured participants' leg length (distance between superior border of the femur's *Greater Trochanter* and the floor), and collected their reported age and gender. The participant watched video instructions which explained the exergame and provided the participant the quitting instruction (https://youtu.be/moWMdY0iRFs).

Next, the participants familiarized themselves with the exergame controls. During this familiarization period, participants played the exergame using the settings of the Standard condition until they felt comfortable playing and after having hit at least thirty moles. The median time taken to hit a mole in the final eight repetitions served as the starting t_{mole} in the Dosed Failure condition. We used the median rather than the mean because the median is more robust to outliers.

Following the familiarization period, we recorded the number of repetitions, play duration, and the 3D position of the feet, while the participant performed each of the conditions of the exergame. The experimenter instructed the participant to play as long as they wanted, and explained that it did not matter for the research how long they would continue. Participants verbally informed the experimenter when they were done playing. As such, the experimenter stopped the exergame manually. If the participant initiated a conversation during the exergame, the experimenter ended the conservation. To avoid excessive fatigue, the exergame was automatically stopped if the participant continued playing for 6 min (twice the playing time), and a short animation was shown to indicate the end of the exergame. Responses to participants' questions during the exergame were standardized and listed in the Appendix. During the exergame, the experimenter stood near the participant to assist in case of a balance loss.

After a participant finished the first condition of the exergame (depending on order group, "Standard" or "Dosed Failure"), they could take a break from the exergame such that they could recover from excessive fatigue. During this break, we assessed self-reported motivation and physical fitness. We assessed his or her subjective motivation with a Quick Motivation Index (QMI; van der Kooij et al., 2019). The QMI is based on the participant's responses to the following two questions:

a. On a scale of 1–10, how much did you enjoy the task until now?

b. On a scale of 1–10, how motivated are you to continue?

After having assessed the QMI, we conducted the FES-i and recorded the participant's maximal handgrip strength by assessing the handgrip strength on three trials for both hands using a handheld dynamometer. The maximum value of the six assessments defined the participant's handgrip strength.

Subsequently, the participant performed the second condition, after which we again assessed their subjective motivation using the QMI. The experiment ended with a short debriefing interview.

Data Analysis

As our primary (objective) measure of motivation, we determined the total effort that the participant put into the exergame as the number of repetitions that were voluntarily played within each condition. Once the participant expressed the wish to quit the exergame, it took a few seconds to manually abort it and during this time moles could be presented. To evaluate the exact moment of quitting, we evaluated whether one of the feet moved toward the mole by determining the direction of velocity. If one of the feet had a velocity above 0.5 m/s in the direction of the mole (a conservative threshold suitable for the noisy signal of the Kinect), we considered this as an attempt. The last time the participant attempted to hit the mole was considered the last repetition before quitting. The number of repetitions in a condition was the number of moles presented until this moment.

In addition to the number of repetitions (our objective measure of motivation), we determined the total play duration in seconds, as well as the mean rating on the two questions in the QMI, reflecting the subjective motivation (see deviation from preregistration for interpretation of these values). To check whether we managed to present about 30% failure, we calculated the average success frequency for each participant by dividing the successful repetitions by the total number of repetitions within a condition.

Sampling Plan and Statistical Analysis

As the effect size of our primary analysis was hard to formulate a priori, a power analysis would have been deceiving. Hence, we did not fix our sample size beforehand and sequentially monitored the Bayes factor (BF) instead. Our main hypothesis was that the dosed failure would lead to a higher number of repetitions played, suggesting a higher motivation for the Dosed Failure condition. This hypothesis was tested using a Bayesian one-sided paired samples t test (Ly et al., 2016) with a Cauchy prior with a scale parameter of $1/\sqrt{2}$. The BF was monitored after each addition of the two new participants from the 20th participant onward. We stopped the experimentation when the BF reached a value of 1/8 or 8 (indicating that one hypothesis is eight times more likely than its alternative), or when a total sample size of 75 was reached. We determined the maximum sample size of 75 participants based on an expected effect size of 1/3 and an alpha set to .05.

The possibility exists that the number of repetitions played was influenced by the order of condition. Therefore, we tested whether our within-subjects design caused order effects in the number of repetitions played, using a Bayesian version of a repeated-measures analysis of variance (ANOVA; Rouder et al., 2012) with conditions (i.e., Standard, Dosed Failure) as within-subject variable and the condition order as a between-subject variable.

For all statistical testing, we checked the model assumptions (i.e., for t test: Shapiro–Wilk's test of normality and visual inspection of the Q–Q plots of the paired differences; for ANOVA: visual inspection of the residuals Q–Q plot, Levene's test of equality of variance). If an assumption was violated, we selected a Bayesian nonparametric equivalent of the statistical test. For the paired t test, a Wilcoxon signed-rank test would have been selected and regarding the ANOVA, a Friedman test would have been used. The median and interquartile range (IQR) were reported in the case of a nonnormally distributed variable.

Data processing was performed in Python and the statistical analysis was done in JASP (version 0.16.1, JASP, 2022). Data and analysis scripts can be viewed in and downloaded from the OpenScienceFramework project page (https://osf.io/azcd6/). Summary statistics were represented as mean ± *SD*. Effect sizes are calculated for statistical tests (Cohen's d for Bayesian paired samples t test, and partial eta squared η^2 for Bayesian repeated-measures ANOVA). Other data, such as gender, reported age, handgrip strength, and the Dutch version of the FES-i, were used to describe our sample characteristics.

Deviations From Preregistration

This article deviates on three aspects from our preregistration. First, we planned to quit experimenting as soon as we reached a BF of 8; however, at that time, we had already scheduled five more participants. We decided to include those participants as well.

Second, we altered the method of determining the starting t_{mole} after the first participant (whose data were excluded from the analysis). According to our preregistered methods, we instructed this participant to hit five moles as soon as possible. The mean time to hit those five moles would serve as the starting t_{mole}. Counter to our expectations, the success frequency did not converge to our proposed 0.5–0.9 bandwidth in the Dosed Failure condition. Hence, we decided to alter the method for all the subsequent participants such that the median time at the final eight repetitions in the familiarization period defined the starting t_{mole}. We thereby omitted the instruction that participants needed to hit the moles as fast as possible.

Third, in the preregistration, we planned a secondary test of our main hypothesis that the motivation in the Standard condition was less than in the Dosed Failure condition on the (1) total play duration and (2) the subjective motivation (QMI). We turned to Bayesian paired sample *t* test instead of Bayesian repeated-measures ANOVAs regarding the testing of the play duration and the subjective motivation (QMI). Residuals of the subjective motivation scores were not normally distributed. To be consistent with our other methods and to improve the interpretability, we decided to test these secondary analyses using the aforementioned test (i.e., Bayesian paired sample *t* test).

Results

In the Dosed Failure condition, we paced the exergame depending on the participant's performance; the median t_{mole} ranged between .8 and 1.6 s/mol. This individual pacing successfully caused failure to hit the mole on about 30% of the repetitions; the success frequency in this condition was on average 0.71 ± 0.06. Participants' fastest t_{mole}, for which they were still successful, had a median of 1.0 (IQR = 0.23). Seven participants kept playing until they hit the 6-min mark in both conditions.

The paired differences of repetitions between conditions seemed normally distributed. The Bayesian paired *t* test showed strong evidence that older adults played more repetitions in the Dosed Failure condition ($n = 164 \pm 75$) compared with the Standard condition ($n = 144 \pm 74$). The evidence for this difference (on an average 14% increase in repetitions) was strong (median Cohen's $d = 0.44$, $BF_{10} = 12.6$; Figure 2A). This interpretation was robust against large changes in the prior distribution (e.g., $BF_{10} = 8.1$ for a wide prior with an *r* scale $\sqrt{2}$).

The effect of the condition appears larger for the group that performed the Dosed failure condition first (lines connecting diamonds steeper than lines connecting disks in Figure 2A). To investigate whether there was an effect of condition order (Figure 2B) on the number of repetitions played, we performed a Bayesian repeated-measures ANOVA; its assumptions were not violated. For both condition order groups, a higher average number of repetitions were played in the Dosed Failure condition (Standard condition first: $n = 138.1 \pm 69.2$, Dosed Failure condition first: $n = 193.7 \pm 72.6$) compared with the number of repetitions played in the Standard condition (Standard condition first: $n = 123.1 \pm 72.6$, Dosed Failure condition first: $n = 166.6 \pm 71.5$). The analysis revealed that the model containing the main effects of the condition and the condition order had the best support from the data ($\eta^2 = .020$, $BF_{10} = 14.3$), compared with the other models. In this favored model, the parameter estimates showed that, in line with our prediction, in the Dosed Failure condition, participants played an

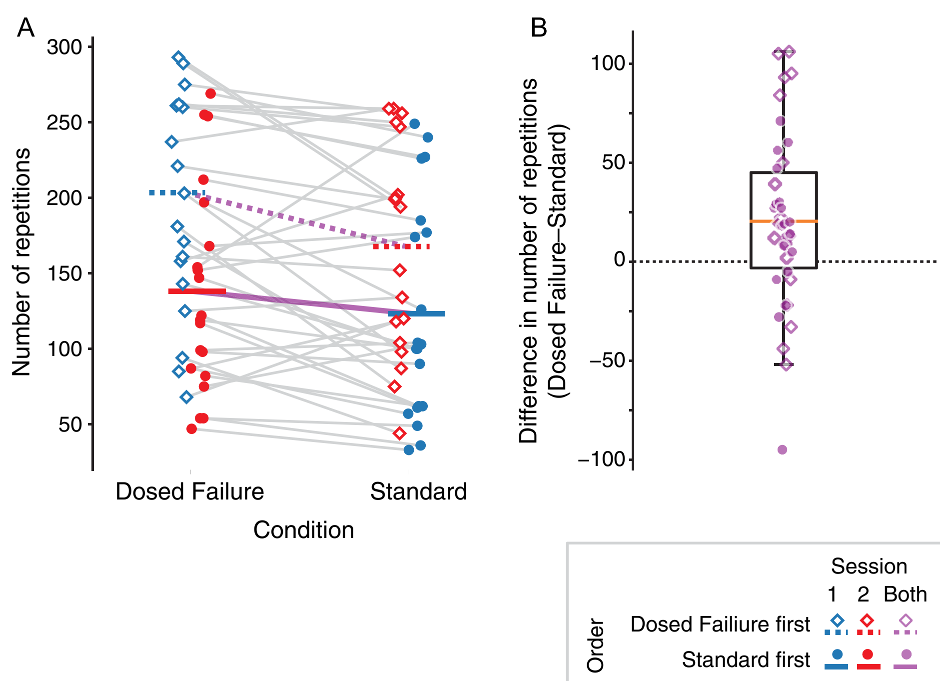

Figure 2 — (A) Scatter plot of the number of repetitions for each participant in the two conditions. The lines connect the data points that belong to the same participant. Participants that started with the Dosed Failure condition are depicted by diamonds, those that started with the Standard condition by circles. The blue (lighter color) and red (darker color) data points show the number of repetitions in the first session and second session, respectively. The horizontal bars indicate the group mean. (B) Box plot of the difference in number of repetitions between conditions. Positive values indicate that the number of repetitions in the Dosed Failure condition is higher than the number of repetitions in the Standard condition. The overlaying dots depict the individual data points. Participants that started with the Dosed Failure condition are depicted by diamonds, those that started with the Standard condition by circles.

additional 19 repetitions more than in the Standard condition. Moreover, the participants in the Standard-first group played on an average 34 repetitions less than the ones in the Dosed Failure-first group. The model containing the interaction between condition and condition order was less likely than the model containing the fixed-effects model without interaction and had a BF_{10} of 5, which is considered not significant.

Regarding play duration, the play duration within the condition was not normally distributed. Therefore, we report the median and IQR for each condition rather than the mean and *SD*. However, the paired differences followed a normal distribution, as shown by a visual inspection of the Q–Q plot and results of the Shapiro–Wilk's test of normality ($W = 0.97$, $p = .31$). As the assumptions for the paired *t* test are that the paired differences are normally distributed, a parametric test was used. Hence, a Bayesian paired samples *t* test was performed and showed the data were inconclusive concerning play duration (Cohen's $d = 0.20$, $BF_{10} = 0.56$). Median play duration in the Dosed Failure condition was 209 s (IQR = 184 s) and median play duration in the Standard condition was 206 s (IQR = 182 s).

The paired differences (see Figure 3A) of subjective motivation were not normally distributed, as indicated by the Shapiro–Wilk's test of normality ($W = 0.81$, $p < .001$). A Bayesian signed-rank test showed that the data (median QMI dosed failure = 8.3, IQR = 1.5, median QMI no failure = 8.0, IQR = 1.5) were inconclusive regarding the hypothesis that ratings on the Dosed Failure condition would be higher than those on the Standard condition ($W = 75$, $BF_{10} = 0.55$).

Additional Analyses

In addition to the preregistered analyses above, we performed two additional tests. First, to explore whether the participants used the score (the displayed number of moles hit) as a quit criterion, we tested whether the two conditions differed in score at the moment the participant quit. If participants decided to quit the second condition when they had hit the same game score, we would expect that the game score did not differ between the two conditions. A Bayesian paired samples *t* test was performed on the game score using a default prior ($r\text{-scale} = 1/\sqrt{2}$), as the paired differences seemed normally distributed. The analysis (see Figure 3B) showed the score in the Dosed Failure condition (115 ± 51) was lower than that in the Standard condition (144 ± 74, Cohen's $d = -0.61$, $BF_{10} = 90.6$).

The second additional analysis is based on the fact that we maximized the playing time to 6 min. Seven out of the 38 participants reached the maximum playing time of 6 min in both the conditions. For these participants, the number of repetitions is thus, not a good measure of motivation. Therefore, we ran an additional analysis on the number of repetitions played in the two conditions without these participants. After the exclusion of these participants, the effect size of the condition on the mean number of repetitions is not reduced (Standard condition: 116, Dosed Failure condition: 139; difference 23; median Cohen's $d = 0.43$). The BF was not reported for this analysis. Data collection relied on the BF, and conducting the analysis on a subset of the available data by default reduces the BF and might potentially lead to misinterpretation.

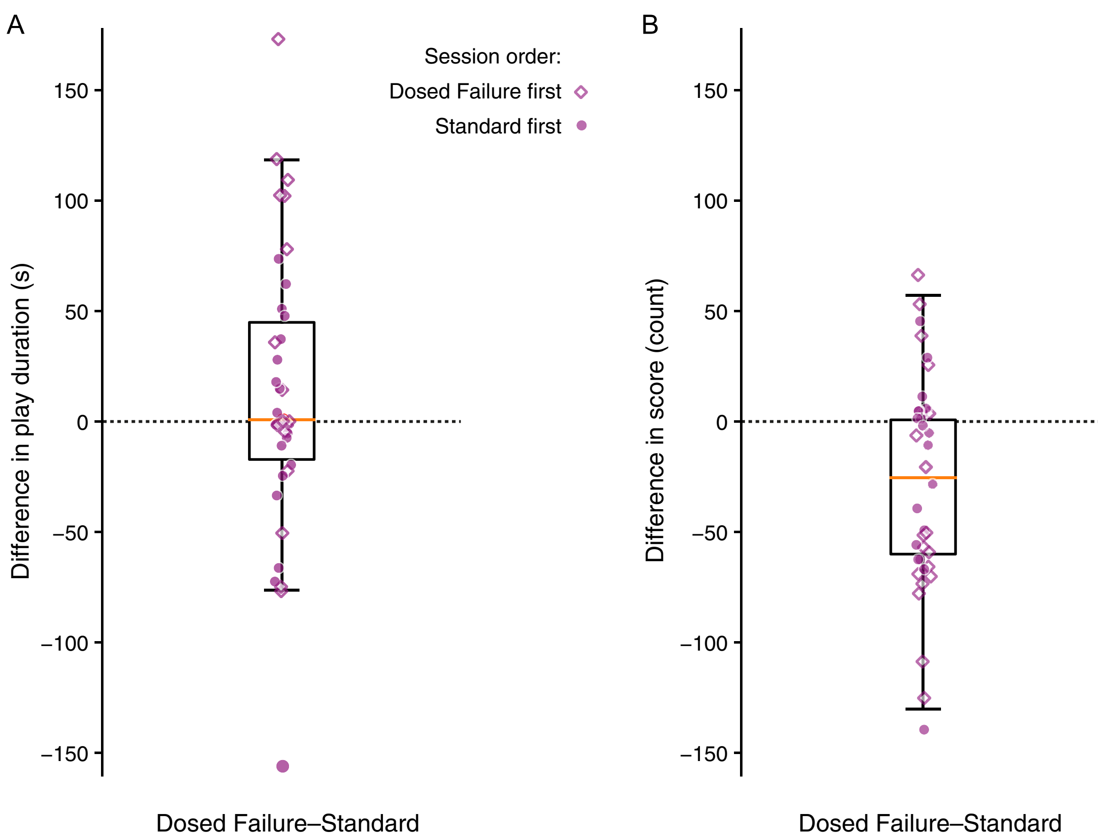

Figure 3 — (A) Difference in play duration between the Dosed Failure condition and the Standard condition. The positive values indicating that the participant played longer in the Dosed Failure condition. (B) Difference in cumulative score between the conditions. Positive values indicate that the individual had a higher score in the Dosed Failure condition compared with the Standard condition. Participants that started with the Dosed Failure condition are depicted by diamonds, those that started with the Standard condition by circles.

Discussion

In this study, we tested the hypothesis that dosed failure increases older adults' motivation for an exergame. In the Dosed Failure condition, the exergame was paced leading participants to fail to hit about 30% of the moles in time, whereas in the Standard condition, there was no pacing and participants hit all the moles. Motivation was measured objectively from the number of repetitions voluntarily played, capturing the total amount of effort the participant had put into the exergame. The results confirmed the hypothesis: participants played more repetitions in the Dosed Failure condition compared with the Standard condition. This was not due to the maximum play duration of 6 min preventing participants to make the repetitions they were motivated to do. When we excluded participants who reached the maximum play duration, the results suggested the same: participants made more repetitions in the Dosed Failure condition. We found no evidence that the play duration or subjective motivation differed between conditions.

Before providing a theoretical interpretation of our results, we first assessed whether the difference in the number of repetitions played between the conditions was caused by the experimental design rather than by a difference in motivation. In our design, we assumed the participant's decision to quit would reflect the participant's motivation. However, one could also imagine that the participant decided to play both conditions of the exergame for the same amount of time or until they reached the same score. Although the play duration did not differ between conditions, it is unlikely that participants set a time-based goal as there was no clock. The finding that the score when participants quit was considerably lower for the Dosed Failure condition argues against the idea that participants played the second session until they achieved the same score as in the first session. Moreover, if participants used a score-based quit criterion, we would have found an interaction of order and condition on the number of repetitions played. That is, participants who played the Standard condition first would set a difficult target (high number of repetitions) for quitting in the Dosed Failure condition, whereas participants who played the Dosed Failure condition first would set an easy target (low number of repetitions) for quitting in the Standard condition. However, we found no evidence that adding such an interaction explained the number of repetitions played better than a model with only the main effects of condition and condition order.

While participants made more repetitions in the Dosed Failure condition, we found no evidence that they played this condition for a longer time than the Standard condition. This indicates that although participants were motivated to invest more energy in the task, the enhanced motivation was not enough to also make them play longer. This means that when using dosed failure to motivate someone, it is important to consider the purpose of the task. If the purpose is physical fitness, introducing failure by pacing the exergame might be wise. If the purpose relies on prolonged engagement, it might be better to introduce failure using accuracy demands which do not require physical effort. We also found no evidence that the dosed failure increased the subjective motivation. This is partially explained by the subjective motivation being measured after participants had decided to quit the exergame. At this moment, the motivation might be expected to be comparable for the two conditions. In addition, the self-reported, subjective motivation, is a less sensitive measure of motivation than the number of repetitions played. Success frequency is only one of the factors impacting the motivation, besides factors such as the social experience of participating in a research study and the novelty of playing an exergame for the first time. Therefore, the effects of success frequency are small, and a sensitive measure of motivation might be required to detect the effect of dosed failure.

Our main contribution to the literature is that we replicated the motivating effect of intermediate challenge in older adults (Beik & Fazeli, 2021; Beik et al., 2022) in a task requiring physical effort, and measuring motivation using an objective measure of effort invested in the task. The finding that dosed failure is motivating is consistent with Flow Theory's prediction that intermediate challenge is optimally motivating (Csikszentmihalyi, 1990). This finding is furthermore consistent with Achievement Motivation Theory's prediction that the relationship between motivation and success frequency follows an inverted-U (Atkinson, 1957). As we operationalize challenge with the success frequency, we also add to the experimental results supporting the inverted-U relationship between motivation and success frequency in young adults (Murayama et al., 2019; van der Kooij et al., 2018, 2021).

Generalization and Study Limitations

Generalization of our conclusion might be limited to relatively fun activities, performed by fit individuals, on a short timescale. Also, although there was no indication that participants decided to play both versions of the game for the same amount of time, we cannot rule out the alternative explanation that instead of using a motivation-based quit criterion, participants used a time-based quit criterion which would have led them to perform more repetitions in the Dosed Failure condition because they were provided a limited time to hit the target.

Subjective motivation was high for both conditions: about eight on a scale of 1–10. For most participants, performing the mole exergame was their first experience with an exergame. It has been pointed out before that this might positively skew the findings on the motivational benefits of exergames (Hamari et al., 2014). In our study, this novelty effect held for both versions of the exergame, but the pleasurable experiences might have especially increased tolerance to experiences of failure.

Although we measured a relatively old group (72 ± 5 years old), participants were mentally (minimental state examination = 29 ± 2) and physically fit. Physical fitness was indicated by the participants' confidence that their daily activities would not result in a fall (FES-i = 21 ± 4), and by handgrip strength. Despite a prevalence of sarcopenia in 70–79 years of 15% (Dodds et al., 2014; Therakomen et al., 2020); only one participant (2.6%) had a handgrip strength that was below the diagnostic threshold for sarcopenia (294 N for men, 196 N for women; Cruz-Jentoft, 2010). It might be that frailer individuals are less motivated by dosed failure because they are more driven to avoid failure, which would result in a preference for a higher success frequency (McClelland et al., 1953). The exergame used has originally been designed for a frailer target group, receiving physiotherapy or living in care homes. The finding that dosed failure increased motivation might be especially relevant to using this exergame in relatively fit target groups. Finally, the generalization to timescales longer than the 3–6 min we measured might be limited.

Interesting avenues for future research are to test the effect of changes in the success frequency on motivation and to test the effect of challenge on learning. A recent "Predictive Processing" theory proposes that humans enjoy success prediction errors—performing better than expected—and are motivated to reduce these prediction errors (Deterding et al., 2022). As a constant success frequency signals that performance does not change (Cowley et al., 2019; Steels, 2004), a constant success frequency might not be optimally motivating. It might therefore be motivating to adapt difficulty to the success frequency in a stepwise design that provides constant pacing for a prolonged episode. Finally, before the results on the motivational effects of challenge can be implemented in the games that are aimed at learning, it should be assessed how the success frequency affects learning. Interestingly, it has been proposed that an intermediate challenge is also optimal for learning (Guadagnoli & Lee, 2004).

Conclusion

To conclude, older adults can be motivated to put more effort into a brief exergame by pacing the exergame such that they fail on about 30% of their attempts.

References

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated goal-directed activities. *Personality and Social Psychology Bulletin, 38*(3), 317–330. https://doi.org/10.1177/0146167211427147

Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6), 359–372. https://doi.org/10.1037/h0043445

Beik, M., & Fazeli, D. (2021). The effect of learner-adapted practice schedule and task similarity on motivation and motor learning in older adults. *Psychology of Sport & Exercise, 54*, Article 101911.

Beik, M., Taheri, H., Kakhki, A.S., Ghoshuni, M., & Fazeli, D. (2022). Contextual interference effects on approach motivation when learning timing tasks: A frontal electroencephalography (EEG) alpha asymmetry study in older adults. *Perceptual and Motor Skills, 129*(4), 1321–1341. https://doi.org/10.1177/00315125221098325

Brehm, J.W., & Self, E.A. (1989). The intensity of motivation. *Annual Reviews in Psychology, 40*, 109–131.

Burgers, C., Eden, A., van Engelenburg, M.D., & Buningh, S. (2015). How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior, 48*, 94–103. https://doi.org/10.1016/j.chb.2015.01.038

Burton, D., & Raedeke, T.D. (2008). *Sport psychology for coaches*. Human Kinetics.

Cook, D.A., & Artino, A.R. (2016). Motivation to learn: An overview of contemporary theories. *Medical Education, 50*(10), 997–1014. https://doi.org/10.1111/medu.13074

Cowley, B.U., Palomäki, J., Tammi, T., Frantsi, R., Inkilä, V.P., Lehtonen, N., Pölönen, P., Vepsäläinen, J., & Lappi, O. (2019). Flow experiences during visuomotor skill acquisition reflect deviation from a power-law learning curve, but not overall level of skill. *Frontiers in Psychology, 10*, Article 1126. https://doi.org/10.3389/fpsyg.2019.01126

Cruz-Jentoft, A.J. (2010). Sarcopenia: European consensus on definition and diagnosis: Report of the European working group on sarcopenia in older people. *Age and Ageing, 39*(4):412–423. https://doi.org/10.1093/ageing/afq034

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Deterding, S., Andersen, M.M., Kiverstein, J., & Miller, M. (2022). Mastering uncertainty: A predictive processing account of enjoying uncertain success in video game play. *Frontiers in Psychology, 13*, Article 924953, https://doi.org/10.3389/fpsyg.2022.924953

Dodds, R.M., Syddall, H.E., Cooper, R., Benzeval, M., Deary, I.J., Dennison, E.M., Der, G., Gale, C.R., Inskip, H.M., Jagger, C., Kirkwood, T.B., Lawlor, D.A., Robinson, S.M., Starr, J.M., Steptoe, A., Tilling, K., Kuh, D., Cooper, C., & Sayer, A.A. (2014). Grip strength across the life course: Normative data from twelve British studies. *PLoS One, 9*(12), Article e113637. https://doi.org/10.1371/journal.pone.0113637

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fong, C.J., Zaleski, D.J., & Leach, J.K. (2015). The challenge-skill balance and antecedents of flow: A meta-analytic investigation. *The Journal of Positive Psychology, 10*(5), 425–446. https://doi.org/10.1080/17439760.2014.967799

Guadagnoli, M.A., & Lee, T.D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior, 36*(2), 212–224. https://doi.org/10.3200/JMBR.36.2.212-224

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?—A literature review of empirical studies on gamification* [Conference session]. Hawaii International Conference on System Science, Waikoloa, HI, USA.

JASP Team (2022). *JASP* (Version 0.16.1) [Computer software].

Kempen, G.I., Zijlstra, G.A., & van Haastregt, J.C. (2007). Het meten van angst om te vallen met de Falls Efficacy Scale-International (FES-I). Achtergrond en psychometrische kenmerken. *Tijdschrift voor Gerontologie en Geriatrie, 38*(4), 178–184. https://doi.org/10.1007/BF03074847

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffrey's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology, 72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. (1953). *The achievement motive*. Appleton-Century-Crofts.

Murayama, K., Sakaki, S., Meliss, S., Yomogida, Y., Matsumori, K., Sugiura, A., Matsumoto, M., & Matsumoto, K. (2019). Motivated for near impossibility: How task type and reward modulates intrinsic motivation and the striatal activation for an extremely difficult task. *bioRxiv*, 828756.

Reijnierse, E.M., de Jong, N., Trappenburg, M.C., Blauw, G.J., Butler-Browne, G., Gapeyeva, H., Hogrel, J.Y., McPhee, J.S., Narici, M.V., Sipilä, S., Stenroth, L., van Lummel, R.C., Pijnappels, M., Meskers, C.G.M., & Maier, A.B. (2017). Assessment of maximal handgrip strength: How many attempts are needed? *Journal of Cachexia, Sarcopenia and Muscle, 8*(3), 466–474. https://doi.org/10.1002/jcsm.12181

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. Cochran-Smith, S. Feiman-Nemser, & D.J. McIntyre (Eds.), *Handbook of research on teaching* (pp. 376–391). Routledge/Taylor & Francis group and the Association of Teacher Educators.

Rouder, J.N., Morey, R.D., Speckman, P.L., & Province, M.P. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ryan, R.M., & Deci, E.L. (2002). *Handbook of self-determination research*. The University of Rochester Press.

Steels, L. (2004). The autotelic principle. In F. Lida, R. Pfeifer, L. Steels, & Y. Kuniyoshi (Eds.), *Embodied artificial intelligence* (pp. 231–242). Springer.

Therakomen, V., Petchlorlian, A., & Lakananurak, N. (2020). Prevalence and risk factors of primary sarcopenia in community-dwelling outpatient elderly: A cross-sectional study. *Scientific Reports, 10*(1), Article 19551. https://doi.org/10.1038/s41598-020-75250-y van der Kooij, K., In 't Veld, L., & Hennink, T. (2021). Motivation as a function of success frequency. *Motivation and Emotion, 45*(6), 759–768. https://doi.org/10.1007/s11031-021-09904-3

van der Kooij, K., Oostwoud Wijdenes, L., Rigterink, T., Overvliet, K.E., & Smeets, J.B.J. (2018). Reward abundance interferes with error-based learning in a visuomotor adaptation task. *PLoS One, 13*(3), Article e0193002. https://doi.org/10.1371/journal.pone.0193002

van der Kooij, K., van Dijsseldonk, R., van Veen, M., Steenbrink, F., de Weerd, C., & Overvliet, K.E. (2019). Gamification as a sustainable source of enjoyment during balance and gait exercises. *Frontiers in Psychology, 10*, Article 294. https://doi.org/10.3389/fpsyg.2019.00294

World Health Organization. (2020). *WHO guidelines on physical activity and sedentary behaviour*.

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomical Bulletin Review, 23*(5), 1–33.

Yardley, L., Beyer, N., Hauer, K., Kempen, G., Piot-Ziegler, C., & Todd, C. (2005). Development and initial validation of the falls efficacy scale-international (FES-I). *Age and Ageing, 34*(6), 614–619. https://doi.org/10.1093/ageing/afi196

Appendix. Responses to Participant Questions During the Exergame (Translated From Dutch to English)

Question:
"For how long does this continue?"
"For how long do I have to play this?"
Response: "Until you've had enough."

Question:
"Is this enough (already)?"
"Do you have enough (already)?"
"Can I take a break?"
Response: "Yes" (and stop the game)

Question:
"Should I continue for (much) longer?"
"Do you have (about) enough?"
Response: "No, you can quit whenever you want."