Utrecht University Repository

Title	Gamified feedback in adaptive retrieval practice: Points and progress-bars enhance motivation but not learning
Authors	van den Broek, Gesa S.E.; Scholten, Stefanie; van Thuil, Bente; van Rijn, Hedderik; van Gog, Tamara; van der Velde, Maarten
Published in	Computers in Human Behavior
Publication Date	2026-04
Link	https://dspace.library.uu.nl/handle/1874/479809
Citation	van den Broek, G S E, Scholten, S, van Thuil, B, van Rijn, H, van Gog, T & van der Velde, M 2026, 'Gamified feedback in adaptive retrieval practice : Points and progress-bars enhance motivation but not learning', Computers in Human Behavior, vol. 177, 108862. https://doi.org/10.1016/j.chb.2025.108862
Versions / License	Publisher version
Rights	https://www.uu.nl/en/university-library/license-and-reuse-conditions

Computers in Human Behavior 177 (2026) 108862



Contents lists available at ScienceDirect

Computers in Human Behavior

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



Gamified feedback in adaptive retrieval practice: Points and progress-bars enhance motivation but not learning

Gesa S. E. van den Broek [a,*], Stefanie Scholten [a], Bente van Thuil [a], Hedderik van Rijn [b,c], Tamara van Gog [a], Maarten van der Velde [b,c]

[a] Utrecht University, Department of Education, the Netherlands
[b] MemoryLab, the Netherlands
[c] University of Groningen, Department of Experimental Psychology, the Netherlands

ARTICLE INFO

Handling Editor: Matthieu Guitton

Keywords:
Gamification
Retrieval practice
Adaptive learning technology
Progress feedback
Motivation

ABSTRACT

Retrieval practice is an effective but effortful learning strategy during which learners bring information from memory to mind (e.g., when self-testing with flashcard software). Because it is effortful, learners often find it hard to maintain during self-study. We therefore investigated if gamified feedback could make retrieval practice more attractive to learners, while maintaining or increasing its benefits for learning. In an online experiment, 166 adult participants studied foreign vocabulary (i.e., practiced retrieval of the word meaning) with an adaptive fact learning system. Each participant practiced once in a control condition, and once in a gamified condition: either a *points* condition which showed a score that increased for every correct answer, or a *points-and-progress* condition, that additionally displayed on two progress bars how the score increased. Compared to the control condition, the gamified conditions enhanced self-reported feelings of competence, enjoyment, and task value as well as experienced goal setting and the desire to perform well. Differences between points and points-and-progress condition were only found for goal-setting (which was higher for points-and-progress). Gamified feedback did not affect learning behaviors during practice, nor recall on a delayed posttest. However, the response speed, score achieved, number of practiced words, and number of trials completed with gamified feedback were lower when participants first practiced with gamified feedback and then without, compared to when they first completed the control condition and then received gamified feedback. Taken together, results suggest that gamified feedback may be a useful addition to enhance motivation for learning with retrieval practice software.

1. Introduction

Retrieval practice is a well-investigated, effective learning strategy during which learners recall information from memory, for example, through self-testing with flashcard software. Compared to other study strategies such as rereading materials, retrieval practice has robust benefits for retention that have been replicated in different groups of learners, with different materials, and in different settings (overviews, for example, in Adesope et al., 2017; Karpicke, 2017; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Rowland, 2014). Adaptive technology can further enhance the benefits of retrieval practice by adjusting practice to individual learners' performance level and to the difficulty of materials (e.g., Pavlik & Anderson, 2008; Sense et al., 2016; van der Velde, Sense, Borst, & van Rijn, 2021). Such technology can, for instance, ensure that difficult items are repeated more frequently and easier items less frequently. In this way, it is ensured that learners need relatively high retrieval effort during practice but have a high chance to retrieve the correct answer successfully, which are characteristics of effective retrieval exercises (Pyc & Rawson, 2009; Rowland, 2014).

Although computerized retrieval exercises are well-suited for self-study, they are at present under-used by many students (Carpenter et al., 2020; Rivers, 2020; but see Weissgerber & Rummer, 2023). A possible reason for this is lack of insight into the benefits of retrieval practice: Students and teachers tend to regard self-testing or retrieval practice as an instrument to monitor rather than enhance learning (for a review, see Rivers, 2020), and learners may misinterpret the higher cognitive effort during retrieval practice compared to other study strategies (e.g., van den Broek et al., 2016) as an indication of unsuccessful learning (Kirk-Johnson et al., 2019; see also Bjork, 2018; de Bruin et al., 2020). In addition, there are mental costs to maintaining cognitively

* Corresponding author. Heidelberglaan 1, 3584 CS Utrecht, the Netherlands.
E-mail address: g.s.e.vandenbroek@uu.nl (G.S.E. van den Broek).

https://doi.org/10.1016/j.chb.2025.108862
Received 9 January 2025; Received in revised form 3 November 2025; Accepted 15 November 2025
Available online 22 November 2025
0747-5632/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

demanding tasks over time, which may make the (comparably effortful) retrieval practice unattractive for learners (e.g., Chevalier, 2018; Feldon et al., 2019; Inzlicht & Schmeichel, 2012; Kurzban et al., 2013). The question therefore arises if and how retrieval exercises can be made more appealing to learners and easier to maintain over time, while preserving the learning benefits. Such more appealing retrieval exercises could improve the learning experience and motivate learners to practice more, allowing them to benefit more from retrieval practice.

The present study tests if retrieval practice can be made more appealing by adding gamification elements. *Gamification* refers to the "process of applying game elements in a non-game context in order to motivate learner behavior" (Bai et al., 2020, p. 2). The purpose of gamification in education is typically to increase persistence (e.g., duration of learner-controlled practice sessions) or invested effort (e.g., attention during exercises) in order to improve learning outcomes (Bernecker & Ninaus, 2021). Given the enormous popularity of recreational games (e.g., Boyle et al., 2012), it is intuitively appealing to use game design to make intrinsically less appealing tasks, such as effortful learning exercises, more attractive. And indeed, recent overview studies suggest that there may be positive effects of gamification on learning (meta-analyses: Bai et al., 2020; Sailer & Homner, 2020; qualitative reviews: Koivisto & Hamari, 2019; Zainuddin et al., 2020; Seaborn & Fels, 2015; see Krath et al., 2021 for a meta-review). However, many prior gamification studies lacked in scientific rigor, which limits strong conclusions (Dichev & Dicheva, 2017; Sailer & Homner, 2020). Sailer and Homner (2020), for instance, summarized gamification effects on cognitive learning outcomes (e.g., posttest score), motivational outcomes (e.g., self-reports), and behavioral outcomes (e.g., accuracy during practice). When analyses were restricted to studies with high methodological rigor (i.e., randomized experiments), the authors found small, significant benefits of gamification on cognitive learning outcomes ($g = 0.42$, $k = 9$), but effects on motivational ($g = 0.22$, $k = 7$) and behavioral measures ($g = 0.27$, $k = 5$) were not significant, possibly due to the small number of eligible studies.

In addition to issues with methodological rigor, there is at present also insufficient information regarding the effect of specific gamification elements, because most studies manipulate multiple elements at once (Bai et al., 2020; Koivisto & Hamari, 2019). Manipulating multiple gamification elements fits a systemic design approach that considers learners' experiences to result from the interplay of the various components of a system (Deterding, 2015). It allows researchers to explore emergent experiences and behavior, which may be difficult to predict from studying individual elements in isolation. However, when multiple elements are manipulated simultaneously, it is challenging to determine the specific component responsible for any observed effects (e.g., Bernecker & Ninaus, 2021). Value-added research (see Mayer, 2020a) to isolate the effect of specific game elements, is needed but still scarce. Keeping track of progress by assigning points, for example, is typically used in combination with badges and leaderboards (Bai et al., 2020), whereas the unique effect on learning (if existent) is unclear. With the aim of advancing insight into the unique effects of individual gamification elements on retrieval practice, the present study investigated effects of points only and points plus progress bars that visualize the points. Both points and progress bars are widely used gamification elements. We investigated effects of these gamification elements on self-reported motivation, behavior during practice (e.g., response speed and accuracy), as well as performance on a delayed posttest.

1.1. Earlier research on gamified feedback: what is known about points and progress bars

1.1.1. Theoretical arguments

Providing a score or points is one of the most common forms of gamification (e.g., Bai et al., 2020; Koivisto & Hamari, 2019). It builds on the idea that points make progress measurable and visible, and thereby provide immediate feedback on learner performance. Such feedback allows learners to strive towards reaching a particular score, which could encourage effort and persistence. Indeed, a combination of feedback and clear, challenging goals is mentioned in most taxonomies of game elements both in the literature on gamification (e.g., Landers, 2014; Seaborn & Fels, 2015; Wilson et al., 2009) and serious games (i.e., full games designed for learning, e.g., Garris et al., 2002; Wouters et al., 2013). To provide goals, points are typically combined with additional features, such as badges (e.g., Higashi & Schunn, 2020), leaderboards (e.g., Landers et al., 2017; Nebel et al., 2016, 2017), or progress bars (e.g., Mekler et al., 2017). These translate points into something that (supposedly) has more inherent meaning or value. For example, a certain number of points might allow users to add a visually appealing icon to their online profile (badge) or rank them higher than classmates on the list of top scorers (leaderboard). The present study focuses on a combination of points with a simple progress bar, which visualizes performance milestones that learners can reach by engaging in the task (similar to the levels visualized by Mekler et al., 2017).

Positive effects of gamified feedback on motivation and/or learning have been explained with reference to different motivational concepts, in particular learners' need for competence as described in self-determination theory (Ryan & Deci, 2000), and principles of goal-setting theory (Locke & Latham, 2002; see Krath et al., 2021, for a meta-review of theoretical concepts in gamification research). Self-determination theory focuses on the role of three basic needs in human motivation: the need for competence (a sense of efficacy), autonomy (being able to make choices), and relatedness or social connectedness (Ryan & Deci, 2000). It is frequently argued that game elements like points, levels, or badges, appeal to learners' need for competence, and may stimulate their internal motivation by indicating progress or good performance (Krath et al., 2021).

Goal-setting is a second theoretical concept that has been used to predict positive effects of feedback mechanisms in gamification and game-based learning (Krath et al., 2021; Tondello et al., 2018). Goal-setting theory is based on the premise that specific, relevant, and difficult yet attainable goals motivate behavior (Locke & Latham, 2002). Within this theory, goals are defined as "object or aim of an action, for example, to attain a specific standard of proficiency, usually within a specified time limit" (Locke & Latham, 2002, p. 705). According to the theory, goals direct attention and effort toward goal-relevant behaviors, increase persistence, and make it more likely that people activate and use relevant knowledge and skills. Gamified feedback like progress bars, can show learners how they progress toward interim-goals and thereby introduce achievements that learners can strive for (e.g., reaching a certain number of points). This might motivate learners in a similar way as explicit goal-setting (or even more effectively, Groening & Binnewies, 2019) and could enhance performance (e.g., the number of responses generated, Landers et al., 2017) and learning.

1.1.2. Empirical studies on the effect of points and progress bars on learning

Although there is anecdotal evidence for possible motivational effects of points during retrieval practice in children (Metcalfe et al., 2009), we did not find any empirical studies on the effect of adding points (or other gamification elements) to retrieval practice on learning outcomes on a posttest. However, there are a few experimental studies on the effect of points on behavioral and motivational outcomes in other tasks. In the context of assessment tests (which involve test questions like retrieval-practice), two experiments showed that adult participants and middle schoolers responded faster yet not more accurately on the test when they received points that rewarded speed of correct answers, compared to a baseline condition without points (Attali & Arieli-Attali, 2015). Middle schoolers also reported higher likability of the test with points. Moreover, adult participants in an online experiment reported being more motivated and feeling more competent, selected more difficult test items, and achieved a higher total score when points were presented (compared to a baseline version of the test, Arieli-Attali, 2016). This was only the case, however, if the point-structure was easy

to understand (rather than complex), did not decrease with incorrect answers, and rewarded test takers for attempting difficult items (Arieli-Attali, 2016). In another study, a combination of points and timer led to significant knowledge gains in a life-support training compared to a baseline condition without points, in which knowledge gains were not significant (Kelle et al., 2013). Presenting only points or only a timer did not have an effect. Finally, providing points per answer increased the quantity but not quality of performance in a repetitive image annotation task, while not affecting self-reported intrinsic motivation and feelings of competence (Mekler et al., 2017). To summarize, the limited prior research on the isolated effect of points supports the idea that points might motivate learners to maintain effort during prolonged tasks. However, more direct empirical evidence is needed regarding effects in the context of retrieval practice, especially regarding the triad of learner motivation, behaviors during practice, and learning outcomes over time.

Although combinations of points with other feedback mechanisms are common in gamification manipulations, we found only two studies that disentangled the effects of different feedback elements experimentally, and no studies with simple visualizations like the progress bars that we used. Mekler et al. (2017) showed that a combination of points with either levels (conceptually these were similar to the progress bars in the present study) or a leaderboard led to a stronger increase in the number of images that participants annotated than points only (compared to the baseline condition without gamification elements). The authors suggested that while points form a connection between user actions and their performance, levels and leaderboards set performance targets for users to aspire to, which drives effects on motivation. In line with this, Chan et al. (2018) found that adding a point system to a mathematics module for adaptive online learning did not have an effect on intrinsic motivation, but a combination of a point system with a leaderboard enhanced motivation. However, there was selective dropout in the different conditions in this study, which prevents strong conclusions. Overall, there are theoretical reasons to predict a positive effect of combining points with additional feedback elements that provide learners with sub-goals that they can strive to reach, but the evidence base is small. Therefore, the present study includes both a simple points condition and a condition that combines points with a bar that visualizes progress.

2. Possible disadvantages of gamified feedback

Whereas the focus in gamification applications is often on positive effects, gamified feedback could also possibly affect learners in an undesired way because the additional elements on the screen may induce extraneous processing (i.e., processing that does not directly support the instructional goal, such as when the learner is distracted by game elements, Mayer, 2020a). Major learning theories like cognitive load theory (Sweller et al., 2011, 2019) and the cognitive theory of multimedia learning (Mayer, 2020b) are based on the premise that human working memory can process only a limited amount of information at once. Therefore, devoting working memory resources to the processing of information that is not directly relevant for learning – such as game-elements on the screen - is thought to limit learners' ability to mentally represent and make sense of the learning materials (Mayer, 2020a, p. 89). While such extraneous processing may be most critical when learners process extensive or complex information (Sweller et al., 2011, 2019), it could also influence retrieval practice: attending to gamification elements could possibly prevent learners from retrieving information from memory, elaborating on materials that were recalled, or from processing feedback presented after failed retrieval attempts (e.g., when they mixed up different items). This, in turn, is likely to hamper learning (see, for example, Hinze & Rapp, 2014, who found that performance pressure – which likely also reduces the available working memory resources – reduced benefits of retrieval practice).

Overall, it is thus not a given that gamified feedback would enhance learning outcomes, as it might also distract learners from the core task.[1] Moreover, distracting effects of gamification manipulations might also interact with adaptive learning systems. For instance, an adaptive learning program that uses reaction times to estimate learners' ability, might need to be re-calibrated to account for slower reaction times caused by learners being distracted by the gamification elements on the screen. Therefore, it is important to measure effects of gamification manipulations on learners' behaviors during practice as well as posttest learning outcomes.

3. The present study

In the present study, we experimentally tested the effects of awarding points for correct answers and visualizing the achieved points with progress bars during retrieval practice of French vocabulary (see Fig. 1). Based on the studies reviewed here before (notably meta-analyses showing positive effects of gamification, Bai et al., 2020; Sailer & Homner, 2020), we predicted that both points and points combined with progress bars would enhance self-reported motivation to engage in (further) practice (Hypothesis 1a), learning behaviors during retrieval practice (Hypothesis 2a), and learning outcomes on a posttest (Hypothesis 3a), in comparison to a control condition with correct-answer feedback and without gamification elements. Moreover, we predicted that the combination of points with progress bars would have a stronger effect than points alone, on motivation (Hypothesis 1b), learning behaviors (Hypothesis 2b), and learning outcomes (Hypothesis 3b), because the progress bars might lend more meaning to the points that learners collect, and provide them with possible (sub-) goals to strive for. Please see Table 1 for an overview of the specific aspects of motivation, learning behaviors, and posttest performance that were measured.

4. Method

4.1. Participants and sample size

Participation in the study was anonymous and voluntary. A link to the experiment was posted on social media and shared by the researchers, inviting native Dutch speakers without knowledge of French to participate. Participants could start the experiment after providing active informed consent. The experiment was started 203 times; 166 participants ($M_{age} = 27.2$ [$sd = 11.4$], 74.6 % female, 25.3 % male, 0 % non-binary; 87/79 in points group/points-and-progress group) completed Session 1 with valid data, and of these, 152 also completed the delayed posttest (82/70 in points group/points-and-progress group; see Appendix I for information on dropout moments and exclusion criteria). The majority of participants in the final sample were students in higher education (57 %) and professionals (27.8 %). The education level was generally high, with 57 % studying towards or holding a university degree, and 31 % studying towards or holding a university of applied sciences degree. Participants were not paid; those who completed both parts of the experiment could enter a raffle to win a gift voucher of 25 euro.

The study was approved by the Ethics Committee of (anonymized),

[1] Note that such a negative effect on learning may be acceptable in situations where positive effects of the distracting game-elements on motivation are important (Mayer, 2020a). For example, in applications that are mostly used during self-study, it may be more important to entice students to practice *at all*, than to develop the most effective exercises by leaving out all game elements.

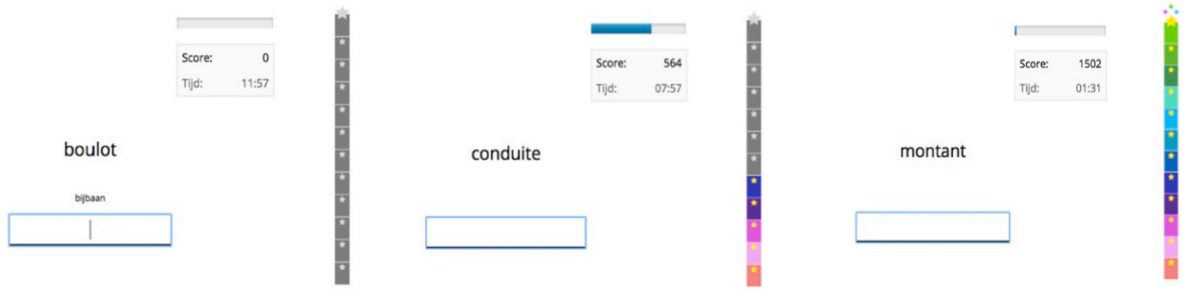


Fig. 1. Gamified Feedback in the Points-and-Progress Condition. *Note:* The figure shows the progress bars during an encoding trial at a score of 0, and during retrieval trials at scores 564 and 1502. The box in the top right displayed a timer counting down from 12 min (present in all conditions), and the current score (gamified conditions only). Additionally, the points-and-progress condition included a colored vertical bar on the right of the screen, which filled incrementally per 100 points achieved; and a horizontal bar that changed with every response and indicated progress towards the next step on the vertical bar (i.e., indicating progress between 0 and 100).

file number 20–695. Given the lack of prior comparable studies, the minimum sample size was determined based on the number of observations required to detect a medium effect size of Cohen's $d = 0.5$ with a power of 0.80, in a between-subject comparison of the two gamified conditions with a *t*-test. This setup requires a sample of at least 51 participants per group, which was reached.

4.2. Design

This study has a 2x2 mixed design, with gamification (control condition versus gamified condition) as within-subject factor and the type of gamification as between-subjects factor (points versus points-and-progress). Assignment of participants to the type of gamification was random. Within participants, the order of conditions was counterbalanced (control first or gamified first), and the assignment of stimuli to conditions was counterbalanced (i.e., there were two matched lists of stimuli; half of the participants studied List A in the control condition and List B in the gamified condition, and vice versa for the other participants).

4.3. Materials

Two lists of each 39 French nouns with Dutch translation were used, such as "nuage = wolk" (English: cloud) and "poignet = pols" (English: wrist). Words were selected from a schoolbook for 15-year old pupils in pre-university education, to make it likely that they would be unknown to learners with limited prior knowledge of French. The two lists contained the same number of words with five, six, seven, eight, or nine characters ($M = 6.6$, $sd = 1.07$). There was variation in orthography in each list (i.e., at least three characters had to be different between all words). All words were checked by a native French speaker with advanced knowledge of Dutch. All stimuli are available at https://edu.nl/bjcbx (OSF repository).

4.3.1. Retrieval practice

In the practice session, every participant completed two blocks of retrieval practice (once in a gamified condition, once in the control condition), followed by questionnaires. We used the *jsPsych* JavaScript library (de Leeuw, 2015) to implement the retrieval practice session and the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) to implement the questionnaires.

Retrieval practice took 12 min per condition. In all three conditions, practice consisted of encoding and retrieval trials with correct-answer feedback. Items that a participant had not seen before were introduced in an *encoding trial*, in which the translation was shown below the French vocabulary and participants copied the translation. In subsequent *retrieval trials*, the French vocabulary was presented on its own and participants were asked to type in the translation from memory, followed by correct-answer feedback that highlighted correct answers in green, and incorrect responses in red while also showing the correct answer. The scheduling of item repetitions and the introduction of new items was determined with an adaptive learning system. This system uses the number and timing of item presentations and the accuracy and speed of learners' responses to distribute the repetition of items during studying, in such a way that items that a learner knows well are repeated less frequently than items that a learner is yet to master (for a detailed description, see Sense & van Rijn, 2022; Sense et al., 2016; van Rijn et al., 2009). Since the system adapts to learners' performance, the number of distinct words encountered during retrieval practice differed between participants (min = 8, max = 39). Accuracy during practice was on average 81 % ($sd = 8.8$ %), which is typical for this scheduling algorithm that is set up to make retrieval effortful but mostly successful (compare, for example, Sense & van Rijn, 2022; van der Velde, Sense, Spijkers, et al., 2021).

4.3.1.1. Descriptors of learning behaviors during retrieval practice. Participants' behavior during retrieval practice was analyzed with respect to the following measures: the accuracy of responses (at trial level, including as correct also responses with minor spelling errors, defined as Damerau-Levenshtein edit distance of 1), reaction times (the onset latency of the first key press after display of the vocabulary prompt), the score achieved at the end of practice (which was visible to learners in the gamified conditions but not visible in the control condition, see next section), the number of words practiced, and the number of trials completed.

4.3.2. Gamification manipulation during retrieval practice: control, points, progress bar

During retrieval practice in the *control condition*, a timer was shown in the top right corner of the screen, which counted downwards from 12:00min. During retrieval practice in the two gamified conditions, additional gamified progress feedback was shown: In the *points condition,* a numerical score was shown above the timer. Starting at 0, the score increased by 10 with every correct response, and reduced by 1 for every incorrect response. This was done to discourage users from skipping difficult items, yet making it likely that the score would increase for the most part, as this was found to be beneficial for motivation in prior research (Arieli-Attali, 2016).

In the *points-and-progress condition*, 2 bars were shown in addition to the timer/score, which visualized the score (see Fig. 1). The maximum score on the progress bar was 1200 points. This score was chosen to offer an achievable but difficult target, based on performance data from an

Table 1
Overview of findings.

Dependent Variable	Outcome Measures	Results		
		Gamified > Control	Progress bar > Points	Other findings
Motivation	Feelings of competence[a]	Yes	No	–
	Enjoyment[a]	Yes	No	Indirect effects: C lower for PP and B2
	Task value[a]	Yes	No	Indirect effects: C lower for PP and B2
	Goal setting[a]	Yes	Yes	Indirect effects: C lower for B2
	Trying to perform well[a]	Yes	Inconclusive (No)	Indirect effects: C lower for B2
	Perceived relevance[a]	Yes	No	No stronger preference for G. with PB than P
Learning behaviors during practice	Accuracy of responses[c]	No	Inconclusive (No)	–
	Reaction times of correct responses[c]	Inconclusive (Yes)	Inconclusive (No)	Int. G x CBO: G faster in B2 than B1
	Total score achieved[a]	No	Inconclusive (No)	Int. G x CBO: G higher in B2 than B1
	Number of trials completed[a]	No	Inconclusive (No)	Int. G x CBO: G higher in B2 than B1
	Number of words practiced[a]	No	Inconclusive (No)	Int. G x CBO: G higher in B2 than B1
Learning outcomes on posttest	Accuracy of responses[b]	No	No	
	Reaction times of correct responses[b]	No	Inconclusive (No)	Int. G x CBO: G faster for items from B2 than B1

Note: This table lists whether evidence for (Yes: $BF_{10} \geq 3$) or against (No: $BF_{01} \geq 3$) effects was found on the different outcome variables for the comparison of gamified and control condition, and for the comparison of points condition and points-and-progress condition. In cases where evidence is inconclusive ($1 < BF_{10} < 3$ or $1 < BF_{01} < 3$) the table indicates the direction in which the anecdotal evidence points. The letters in superscript indicate at which level the outcome was measured. [a]at participant level, per condition. [b]at item level. [c]at trial level. Evidence for or against interactions is mentioned in the third column. A visual summary of evidence strength is provided in Appendix III. Abbreviations: *B1* = first practice block; *B2* = second practice block; *C* = control condition, *P* = points condition, *PP* = points-and-progress condition, *G* = gamified condition (P or PB); *CBO* = counterbalancing order, *Int.* = Interaction effect of, *RT* = reaction time.

earlier, unpublished experiment.[2] To avoid that fast learners would fill the bar before the end of the practice time and then have no further incentive to keep practicing, up to three stars were added at the top of the progress bar when someone reached a score of 1300, 1400 or 1500 points. These stars were not visible while the score was below 1300, and learners were not informed beforehand that the stars would appear, though they did see that the horizontal progress bar continued to fill. There was no further visual feedback when participants reached a score of 1600 or higher; this occurred for 5 out of the 77 participants (6.5 %) in the points-and-progress condition.

4.3.3. Questionnaires to measure motivational outcomes

All items described below were answered on a 7-point Likert-scale, ranging from (1) *not true* to (7) *true*.

4.3.3.1. Intrinsic motivation (enjoyment), feelings of competence, and task value. Eleven items from the Intrinsic Motivation Inventory (IMI, Center for Selfdetermination Theory, n.d.) were used to measure participants' subjective experience of practice in terms of enjoyment (five items of the "intrinsic motivation" scale, e.g. "I enjoyed this exercise."), feelings of competence (four items, e.g., "I am satisfied with my performance during this exercise"), and task value (two items, e.g., "I think that this exercise helps me learn new words").[3] For statistical analyses, the average response was calculated per subscale, resulting in values ranging from 1 to 7. The subscales all had high internal consistency (all scales $\alpha > 0.85$).

4.3.3.2. Goal setting and performance goals. We adapted items regarding goal use and performance goals from previous research on goal setting (Kwan & Lee, 2013) to measure to what extent (i) participants felt like they were working towards a goal during practice (*goal setting*), and (ii) aimed to perform well.

4.3.3.3. Stress and distraction. To measure possible negative outcomes of practice, participants rated to what extent they found practice stressful (another item from Kwan & Lee, 2013) and experienced the gamified feedback as distracting ("*During practice I was distracted by the score/by the score and progress bars.*").

4.3.3.4. Perceived relevance and task preference. Participants rated three items concerning the perceived relevance of the gamified feedback (based on Higashi & Schunn, 2020). These items concerned the clarity of the feedback, the value of adding the feedback to practice, and the extent to which participants tried to achieve a high score. Crombach's alpha for the composite scale was 0.79 (95 % CI = [0.76–0.83]). In addition, participants were asked at the end of the practice session which of the two retrieval tasks they would use to learn words on a different occasion as a direct measure of task preference.

4.3.3.5. Judgments of learning. Participants also rated on a continuous scale (0–100 %) which percentage of the practiced translations they thought they would still know in two days. This item was added as part of a student project. As we did not have hypotheses regarding this outcome measure, we report it in the online data repository (see https://edu.nl/bjcbx).

4.3.4. Delayed posttest to measure learning outcomes

A posttest was administered 2–3 days ($M = 2.19$, $sd = 0.57$) after the practice session to measure learning outcomes. On the posttest, participants were presented with the French words one by one and had to type in the Dutch translation. Spelling mistakes were counted as correct if no more than two editing operations were needed to correct the answer (e.

[2] We analyzed unpublished data from an experiment in which secondary education students (grade 8/9 of high school tracks preparing for higher education) practiced the stimuli of this experiment in the control condition for 12 min. More than 75 percent of the participants in the prior study would have achieved a score that filled at least half of the level bar, but less than 25 % would have reached the full score (25th/50th/75th percentile in prior experiment: 641/790/939).

[3] The IMI subscales "experienced control", "effort" and "experienced pressure" were not used in the present study. The IMI provides reliable measures also when only subscales are selected (Center for Selfdetermination Theory, n.d.).

g., *suces* or *succse* instead of *succes*). We used the stringdist function in R (stringdist_0.9.5.5, method "lv", van der Loo, 2014) to calculate edit distances automatically.

Due to a technical error, the posttest contained only 70 of the 78 words that were practiced (3 words were missing from List 1, and 5 were missing from List 2). Therefore, we measured learning outcomes as the proportion of words that were recalled correctly on the posttest relative to the number of practiced words that were tested (e.g., if someone practiced 19 words in a condition, but only 18 of these words were tested, and the participant translated 9 of these words correctly on the test, then the posttest score was 9/18 = 0.5).

4.3.4.1. Prior knowledge. To obtain an estimate of prior knowledge, we calculated the proportion of unpracticed words that were recalled correctly on the posttest (i.e., those words out of the total number of 70 tested stimuli that had not been included in practice but that were recalled correctly on the test). For example, a participant who practiced 20 words in the control condition, and 30 words in the gamified condition, was tested on 20 + 30 = 50 practiced and 70-50 = 20 unpracticed words. If the participant translated one of these unpracticed words correctly, s/he would have a proportion correct of 1/20 = 0.05 for unpracticed items. Between participants, the number of unpracticed words varied between 2 and 52; for 12 participants who managed to practice all words, no measure of prior knowledge was available. On average, the remaining participants translated 4.3 % of the unpracticed words on the posttest (*sd* = 8.7 %), suggesting low prior knowledge. Four participants had comparably high prior knowledge (above *M*+3sd = 31.7 %). One of them was removed because s/he admitted cheating; the other three were retained in the dataset because some variation in prior knowledge is to be expected in a learner population.

4.4. Procedure

The study consisted of two sessions, in which participants worked online at a desktop computer or laptop (settings prevented users from accessing the experiment via mobile phones or tablets). Session 1 took about 30 min and comprised consent forms and the two practice blocks; Session 2 comprised the posttest and written debriefing and took about 15 min. The link to Session 2 was shared via e-mail after 2 days, and remained available until the end of the next day (M_{delay} = 2.19 days, *sd* = 0.57).

4.5. Transparency and openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The data, analysis code, code for visualizations, and the stimuli are available at https://edu.nl/bjcbx (OSF repository, anonymized for review purposes). Access to the experimental task is available by emailing the corresponding author.

4.6. Analyses

We used mixed-effects regression models (Baayen et al., 2008) to assess the effect of gamified feedback on different measures of (i) participants' self-reported motivation and goal setting behaviors, (ii) learning behaviors during practice, and (iii) learning outcomes on the posttest, while accounting for possible differences between blocks and experimental groups. Each model contained the same set of predictors (see also Appendix II): To determine whether gamification in general affected outcomes, all models included a main effect of *gamification* (*gamified* versus *control condition*). Differences between the two types of gamified feedback were assessed through an interaction term for *gamification* and *experimental group* (*points* or *points-and-progress*). In addition, to control for a possible impact of the counterbalancing order, the models included an interaction term for the within-subjects factor *gamification* (gamification vs. control) and *order* (*gamified-first* or *control-first*). Since this order effect may differ between the two types of gamified feedback, the models also included a three-way interaction between *gamification*, *order*, and between-subjects factor *type of gamification*. To be able to evaluate these different effects separately, we mean-centered the categorical predictors *order*, and *type of gamification* This approach allowed us to draw a general conclusion about the effect of gamification while quantifying differences between specific orders and/or gamification types. The models included a random intercept for participants and, wherever possible, a random intercept for facts.

All dependent variables are listed in Table 1. Response accuracy (a binary outcome) was modelled using a logistic mixed-effects regression model. All other measures were modelled using linear mixed-effects regression. Response times were log-transformed before modelling to improve the normality of the residuals. We used Bayesian methods for the reported analyses, to be able to quantify evidence in favour of the null hypothesis in the absence of an effect (Gallistel, 2009). Regression models were fitted using the *brms* package (version 2.21.0; Bürkner, 2017) for R (version 4.3.1; R Core Team, 2021). We used weakly informative priors for fixed effects that were scaled according to the expected magnitude of coefficients: a Normal(0, 0.1) prior for log-transformed response times, a Normal(0,1) prior for the 7-point motivational outcomes and response accuracy, a Normal(0, 3) prior for the number of words practiced, a Normal(0, 10) prior for the number of trials completed, and a Normal(0, 100) prior for total score achieved. We then evaluated effects through Savage-Dickey density ratio tests on the posterior distributions of each fitted model coefficient β (Wagenmakers et al., 2010). This test quantifies the shift in evidence in favour of or against the null hypothesis (no effect; $\beta = 0$) after having observed the data, by dividing the posterior density at $\beta = 0$ by the prior density at that point. The resulting value is a Bayes factor, BF_{01}, for each predictor in the model. When $BF_{01} > 1$, the data provide evidence in favour of the null hypothesis and against the alternative hypothesis (an effect; $\beta \neq 0$). Conversely, when $BF_{01} < 1$, the data provide evidence against the null hypothesis and in favour of the alternative hypothesis. In this case, we report the inverse Bayes factor to quantify evidence for the effect: $BF_{10} = \frac{1}{BF_{01}}$. When $BF_{10} = 10$, for example, the data are 10 times more likely under the alternative hypothesis that there is a difference between the conditions, than under the null hypothesis that no such difference exists. While Bayes factors express evidence on a continuum, we follow commonly used conventions in interpreting evidence strength (Jeffreys, 1961; Kass & Raftery, 1995; Lee & Wagenmakers, 2014) to describe evidence as "anecdotal" (1 < BF < 3), "moderate" (3 < BF < 10), "strong" (10 < BF < 30), or "decisive" (BF > 30). In addition, we used the *BayesFactor* package for R (Morey et al., 2024) for a Bayesian test of association, with default settings.

5. Results

5.1. RQ1: Does *gamified feedback change* learners' motivation for retrieval practice?

We analyzed effects of the gamification manipulation on several components of learners' motivation (see Table 1 and Appendix III). The complete output of all regression models reported hereafter can be accessed at https://osf.io/tzrxf.

5.1.1. Feelings of competence

We found decisive evidence that self-reported feelings of competence were higher in the two gamified conditions (M_{est} = 4.40 on the seven-point scale) than in the control condition (M_{est} = 3.96, see Fig. 2), b = 0.45, BF_{10} = 827.23, but moderate evidence against a difference in feelings of competence between points condition and points-and-progress condition, b = 0.12, BF_{01} = 3.91. Effects on feelings of competence did not depend on the order in which the conditions were

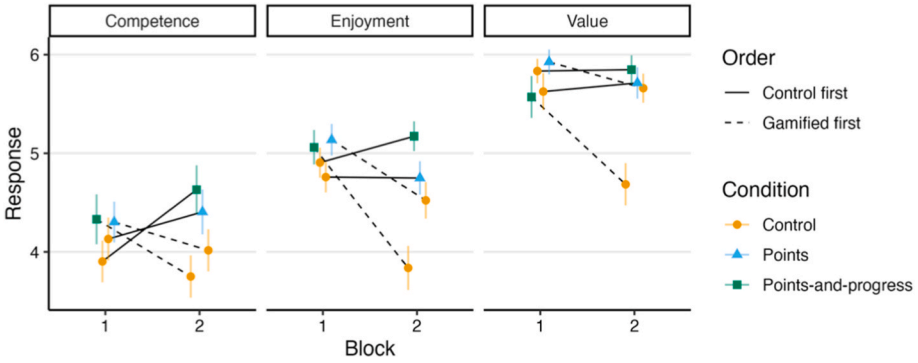


Fig. 2. Self-reported Feelings of Competence, Enjoyment, and Value of Practice for Learning per Condition per Practice Block. *Note:* Self-reports of task motivation were collected after each block of retrieval practice. The plot displays outcomes averaged across participants (+- *SE*), per condition, split by counterbalancing order. Each line represents one group of participants. Data points were horizontally jittered for visualization.

presented (i.e., there was anecdotal to moderate evidence in favour of the null hypothesis for all other interaction terms in the model).

5.1.2. Enjoyment

There was also decisive evidence for enjoyment being higher in the gamified conditions (M_{est} = 5.03) than in the control condition (M_{est} = 4.49), b = 0.54, BF_{10} = 9.87*10^{18}, and moderate evidence against a difference between points condition and points-and-progress condition, b = 0.16, BF_{01} = 3.97. In addition, there was decisive evidence for an indirect effect of gamification in that the control condition was enjoyed less if it was presented *after* the gamified condition (M_{est} = 4.21) rather than *before* the gamified condition (M_{est} = 4.82), b = −0.61, BF_{10} = 125.31, and there was moderate evidence that this effect was more pronounced for the points-and-progress condition (M_{est} = 3.87 versus M_{est} = 4.86) than for the points condition (M_{est} = 4.50 versus M_{est} = 4.77), b = - 0.71, BF_{10} = 3.40.

5.1.3. Task value

The data provide decisive evidence that participants rated the value of retrieval practice for learning higher in the gamified conditions (M_{est} = 5.77) than in the control condition (M_{est} = 5.44), b = 0.33, BF_{10} = 1.29*10^{7}, but moderate evidence against a difference in task value between the points condition and points-and-progress condition, b = −0.12, BF_{01} = 4.72. In addition, there was strong evidence that the value of the control condition was rated lower if it was presented after the gamified condition (M_{est} = 5.20) rather than before the gamified condition (M_{est} = 5.72), b = −0.51, BF_{10} = 14.02, and more so by learners in the points-and-progress group than by learners in the points group (i.e., there was moderate evidence for an interaction indicating that the control condition was rated lower by the learners in the points-and-progress group, b = −0.42, BF_{10} = 3.35, and very strong evidence for a (to be expected) three-way-interaction indicating that this effect was stronger when the gamified condition was presented first, b = −1.04, BF_{10} = 46.33).

5.1.4. Experienced goal setting

We found decisive evidence that, after practice in the gamified conditions, learners reported more that they felt like they were working towards a goal during practice (M_{est} = 5.25) compared to the control condition (M_{est} = 4.35), b = 0.90, BF_{10} = 1.49*10^{18}, with moderate evidence for a difference between the points-and-progress condition (M_{est} = 5.59) and the points condition (M_{est} = 4.96), b = 0.61, BF_{10} = 5.01. In addition, there was very strong evidence that learners reported less goal setting in the control condition when the control condition was presented after gamified practice (M_{est} = 3.99) than before gamified practice (M_{est} = 4.79), b = −0.81, BF_{10} = 34.85, and anecdotal evidence for less goal setting in the control condition within the points-and-progress group (M_{est} = 4.05) than in the points group (M_{est} = 4.60), b = −0.54, BF_{10} = 2.69.

5.1.5. Trying to perform well

We found decisive evidence that learners reported that they tried to perform well more during practice in the gamified conditions (M_{est} = 5.75) than in the control condition (M_{est} = 6.24), b = 0.50, BF_{10} = 6.47*10^{17}. There was anecdotal evidence against a difference between the points and points-and-progress condition, b = 0.24, BF_{01} = 2.04. However, there was again an effect of the manipulation on the control condition: we found decisive evidence that learners tried harder to perform well in the control condition when they started in the control condition (M_{est} = 6.06) than when they first practiced in the gamified condition (M_{est} = 5.48), b = −0.59, BF_{10} = 242.42, and anecdotal evidence that they tried harder to perform well in the control condition when in the points group (M_{est} = 5.92) than in the points-and-progress group (M_{est} = 5.56), b = - 0.36, BF_{10} = 2.17.

5.1.6. Which type of feedback do learners prefer?

There was moderate evidence that participants rated the relevance of the gamified feedback as higher in the points-and-progress condition (M_{est} = 5.36) than in the points condition (M_{est} = 4.76), b = 0.62, BF_{10} = 8.70. However, a Bayesian test of association found moderate evidence that participants in the points-and-progress group did not show a stronger preference for the gamified condition over the control condition than participants in the points group, BF_{01} = 6.41: After practice with points and progress bars, 70.1 % of participants preferred gamified practice over the control condition, while 20.8 % preferred the control condition and 9.1 % had no preference. After practice with only points, 58.6 % preferred the gamified condition, 25.3 % preferred the control condition, and 16.1 % had no preference.

5.1.7. Summary RQ1

Self-reports of task motivation showed both a direct effect of the gamification manipulation - with higher ratings for the gamified conditions than the control condition with regard to feelings of competence, enjoyment, task value, experienced goal setting, and efforts to perform well - and an order effect where most ratings of the control condition were lower when it was presented in Block 2 (after a block of gamified practice), compared to when the control condition was presented in Block 1. This effect was more pronounced in the points-and-progress group for enjoyment and task value ratings. Besides this, we found only one difference between the two gamification conditions, with moderate evidence that learners felt more like they were working towards a goal in the points-and-progress condition than in the points condition. For all other measures of motivation, we found anecdotal to moderate support for the null hypothesis that the two gamification conditions did not differ.

5.2. RQ2: Does gamified feedback change learning behaviors during adaptive retrieval practice?

We analyzed effects of the gamification manipulation on learning behaviors both at trial level (accuracy and speed of practice responses) and at the level of participants (analyzing the total score achieved, the number of trials completed, and the number of words practiced), see Table 1 and Appendix III. Fig. 3 shows an overview of the five learning behaviors per condition, split by counterbalancing order.

5.2.1. Accuracy during practice

We found very strong evidence that response accuracy during practice did not differ between the gamified conditions and the control condition (b = - 0.01, BF_{01} = 37.52), as well as anecdotal evidence against a difference between the two gamification conditions (points versus points-and-progress, b = - 0.16, BF_{01} = 2.70), and anecdotal to moderate evidence against indirect or order effects.

5.2.2. Response times

We only found anecdotal evidence for an overall difference in response times for correct answers between the gamified conditions and the control condition (b = 0.02, BF_{10} = 2.41), and anecdotal evidence against a difference between the two gamification conditions (b = 0.04, BF_{01} = 1.33), which means that the data are largely inconclusive with respect to a main effect of gamification on response speed. There was moderate evidence for an effect of block order on the difference in response times between the gamified conditions and the control condition (b = 0.07, BF_{10} = 4.36): predicted reaction times in Block 1 were $M_{gamified}$ = 1703ms, $M_{control}$ = 1621ms, while predicted reaction times in Block 2 were $M_{gamified}$ = 1589ms, $M_{control}$ = 1614ms. There was moderate evidence against an indirect effect of the type of gamification on reaction times in the control condition (b = 0.01, BF_{01} = 3.49).

5.2.3. Total score, words practiced, trials completed

The Bayesian regression analysis produced moderate evidence against an overall effect of the gamification manipulation on the total score achieved during practice (see Fig. 4), the number of words practiced, or the number of trials completed. That is, the gamified conditions and the control condition did not differ on these measures (total score: b = −4.70, BF_{01} = 5.41; number of words practiced: b = 0.003, BF_{01} = 6.73; number of trials completed: b = −0.81, BF_{01} = 6.05). The data were fairly inconclusive about the two types of gamification being different from each other, with only slight anecdotal evidence against an effect of gamification type (total score: b = −51.82, BF_{01} = 1.07; number of words practiced: b = - 1.29, BF_{01} = 1.39; number of trials completed: b = −3.62, BF_{01} = 1.67). However, practice was more efficient when the gamified feedback condition was presented after the control condition than when it was presented first: learners achieved on average about 103 more points (b = −103.71, BF_{10} = 9.60), practiced about 3 more words (b = −2.57, BF_{10} = 4.00), and completed about 10 more trials (b = −10.34, BF_{10} = 17.47), during gamified practice when it happened in the second block rather than the first block (moderate to strong evidence).

5.3. RQ3: Does gamified feedback change learning outcomes on a delayed posttest?

Posttest data (recall accuracy and reaction times for correct answers) are shown in Fig. 5. We found strong evidence *against* an effect of Gamification on recall accuracy on the posttest (gamified vs. control, b = 0.01, BF_{01} = 18.99), suggesting that recall did not differ between the gamified conditions and the control condition. There was moderate evidence against an effect of Type of Gamification (points vs. points-and-progress, b = −0.17, BF_{01} = 3.54), suggesting that recall did not differ between the two gamification conditions.

We also found moderate evidence *against* a main effect of Gamification on reaction times for correct answers on the posttest (gamified vs. control, b = 0.01, BF_{01} = 5.62) and anecdotal evidence against an effect of Type of Gamification on reaction times (points vs. points-and-progress, b = 0.03, BF_{01} = 1.66). There was moderate evidence for an effect of order: Reaction times for correct responses were about half a second faster on the posttest when words had been presented in the gamified condition in the second practice block (M_{est} = 4165ms) rather than the first practice block (M_{est} = 4561ms), b = 0.09, BF_{10} = 3.69. There was only anecdotal evidence for a difference between the two types of gamification with respect to the effect of order, b = 0.09, BF_{10} = 1.59.

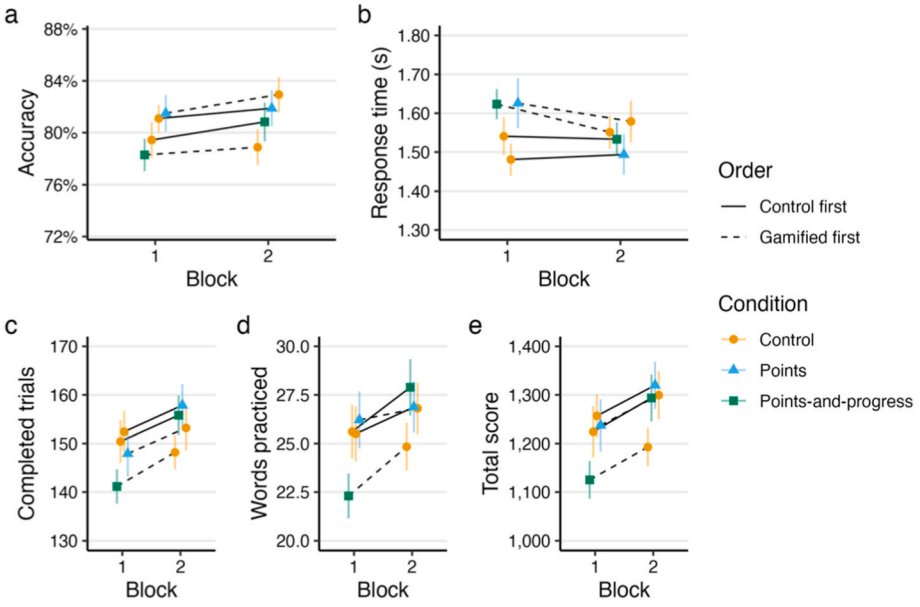


Fig. 3. Learning Behaviors during Adaptive Retrieval Practice with and without Gamified Feedback. *Note:* Plot displays outcomes averaged across participants (+- SE), split by counterbalancing order. Data points were jittered for visualization. For response times, the median was first calculated per participant before aggregating across participants.

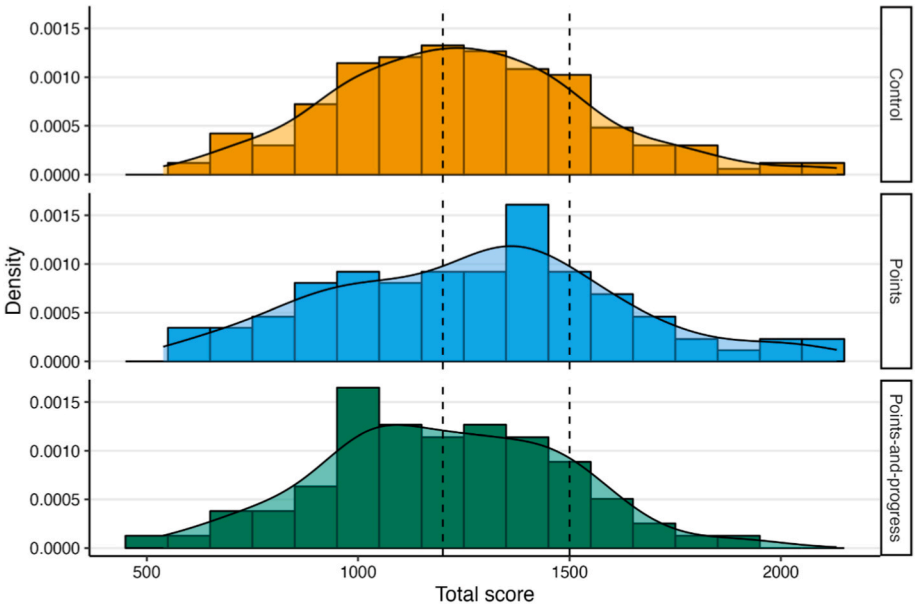


Fig. 4. Distribution of Total Scores that Participants Achieved at the End of Practice, per Condition. *Note:* The score was visible during practice in the points condition and the points-and-progress condition, but not in the control condition. The dotted vertical lines indicate scores of 1200 and 1500. In the points-and-progress condition, the vertical progress bar was completely filled at 1200 points. At 1500 points, all three bonus stars on top of the bar were filled and no further visualization of progress was added afterwards. There was no difference between the total scores achieved in the control condition and the gamified conditions, nor between the two gamified conditions.

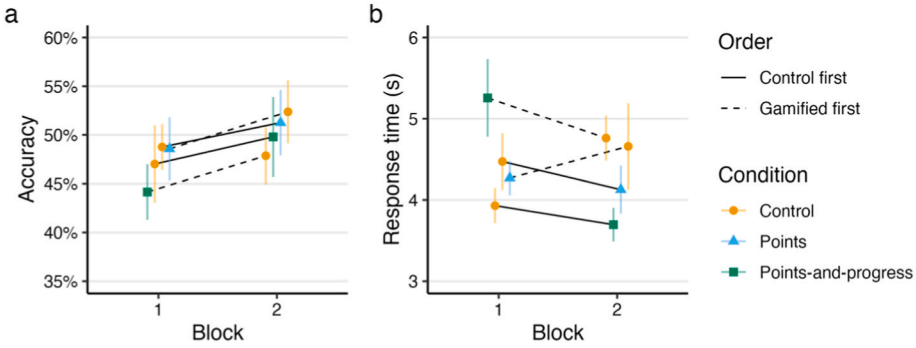


Fig. 5. Recall Accuracy and Response Times on the Delayed Posttest, per Condition, Split by Counterbalancing Order. *Note:* Plots display outcomes averaged across participants *(+- SE)*, split by counterbalancing order. Data points were jittered for visualization. For response times, the median was first calculated per participant before aggregating across participants.

5.4. Exploratory analyses

5.4.1. Does gamified feedback induce stress or distraction?

To explore possible negative outcomes of gamification during retrieval practice, we analyzed participants' rating of the extent to which they found practice stressful and experienced the gamified feedback as distracting. The data suggest that participants experienced gamified practice ($M_{est} = 3.96$) as more stressful than the control condition ($M_{est} = 3.59$), $b = 0.36$, $BF_{10} = 12.47$ (strong evidence), but *not* more distracting than the timer in the control condition, $b = 0.01$, $BF_{01} = 7.00$ (moderate evidence). Moreover, participants experienced the points-and-progress condition ($M_{est} = 4.26$) as more stressful than the points condition ($M_{est} = 3.66$), $b = 0.58$, $BF_{10} = 5.16$ (moderate evidence), and experienced the points-and-progress condition as more distracting, $b = 0.54$, $BF_{10} = 3.08$ (moderate evidence), although both conditions were on average rated relatively low on the 7-point distraction scale ($M_{est} = 3.23$ and $M_{est} = 2.69$, respectively). There was also strong to decisive evidence for an interaction effect with order, indicating that gamified feedback was perceived as more distracting when it was presented after first engaging in the control condition than when gamified feedback was presented first, $b = -0.86$, $BF_{10} = 123.01$, whereas the timer in the control condition was rated as *less* distracting when participants engaged in the control condition after an initial round of practice with gamified feedback, $b = -0.66$, $BF_{10} = 11.94$.

6. Discussion

We investigated the effects of two types of gamified feedback within adaptive retrieval practice, awarding points based on accuracy and visualizing the achieved points with progress bars, in order to test the effect of gamified feedback on learners' task motivation and goal setting, learning behaviors, and learning outcomes on a posttest, in comparison to a control condition without feedback.

6.1. Learners were motivated by gamified feedback (RQ1)

The gamification manipulation enhanced all aspects of self-reported motivation and goal setting behavior. Task motivation (feelings of competence, enjoyment of practice, and task value) as well as goal setting behaviors (feeling like working towards a goal and trying to

perform well) were all rated higher in the gamified conditions than in the control condition. This was in line with Hypothesis 1a. In contrast to Hypothesis 1b, however, none of the measures except for the feeling of working towards a goal differed between the points condition and the points-and-progress condition, with moderate evidence for the null hypothesis.

Different explanations are given in the literature regarding how gamified feedback influences motivation (Krath et al., 2021). For one, feedback is thought to appeal to learners' need for competence by indicating progress or good performance, improving learners' experience of practice, and thereby internal motivation. In another line of reasoning, gamified feedback is said to stimulate goal setting behavior. Our data seem to support both accounts with regard to the overall comparison of gamified and non-gamified practice. For the comparison of the two gamification conditions, the points-and-progress condition only led to more experienced goal setting than the point condition – we assume because the visual bars translated the numerical score into (interim) goals that learners could strive to reach – but did not lead to higher intrinsic motivation than the points condition.

Interestingly, we also found indirect effects where learners found the control condition less motivating when it was presented after practice with gamified feedback than when it was presented in the first practice block. This might reflect a comparison effect, where participants were less motivated to engage in the control condition when they knew that there were also more elaborate, gamified versions of practice. Moreover, the control condition was evaluated less positively in the points-and-progress group than in the points group (in terms of enjoyment and task value). Thus, while the points-and-progress condition itself was not evaluated more positively than the points condition, removing points and visualizations seems to have affected learners' experience of the basic control condition more negatively than removing only the points.

6.2. Gamified feedback in the first round of practice reduced response speed (RQ2)

In the analyses of learning behaviors, we found strong evidence for similar response accuracy in the gamified conditions compared to the control condition. There was no clear gamification effect on reaction times, but reaction times were longer during practice with gamified feedback when the feedback was presented in the first round of practice (whereas after a round of practice in the control condition, this was not the case). In line with these longer reaction times, learners engaging in the gamified condition first also achieved a lower total score, practiced a smaller number of words, and completed a lower number of trials in the gamified conditions, than those who engaged in the gamified condition last. This suggests that the combination of getting acquainted with a new learning task in the first practice block combined with receiving gamified feedback, may have overwhelmed learners and negatively influenced learning behaviors. In contrast, when learners had first gained experience with the task in the control condition, there were no (negative) effects of the gamification manipulation on the learning behaviors.

These findings regarding learning behaviors contradict our prediction (Hypotheses 2a and 2b) that gamified feedback, and in particular the points-and-progress condition, would lead learners to try harder during practice in order to achieve a higher score or reach the next step on the progress bar (e.g., Bai et al., 2020; Koivisto & Hamari, 2019). On the one hand, these results are in line with the idea that gamified feedback, at least when presented from the start, might induce extraneous processing (i.e., processing that does not directly support the instructional goal; Mayer, 2020a) comparable to a kind of seductive detail (Rey, 2012; Sundararajan & Adesope, 2020) that causes learners to devote attention away from the core task. On the other hand, no differences were found between the points and points-and-progress condition, even though the latter included additional, more visually complex information on the screen (and distracted learners more than the simple score according to their self-reports, see results section). More importantly, if there was indeed extraneous processing, it does not seem to have hampered learning outcomes in the present study.

6.3. Few effects of gamified feedback on posttest performance (RQ3)

Contrary to our initial predictions (Hypothesis 3a/b), learning outcomes (i.e., recall performance) on the delayed posttest did not differ between the gamified conditions and the control condition nor between the two types of gamified feedback. That is, based on the present findings, the amount of information that was recalled successfully on a test two to three days after practice was comparable in all three conditions.

However, the gamification manipulation did have some effect on the response speed on the posttest: Words that were translated correctly on the posttest, were translated slower when they had been practiced in the gamified condition in the first practice block rather than the second practice block. This parallels the pattern of results that we found for learning behaviors during practice, suggesting that a negative effect of encountering gamified feedback in an unfamiliar task, may not only influence performance during practice but also translate into later recall speed. Possibly, being able to complete more practice trials allowed learners to also become faster at later recall. However, the size of this effect is relatively small, and results should be replicated in future research before stronger conclusions are drawn.

One possible explanation for the lack of gamification effects on learning outcomes is that the higher motivation with gamified feedback (see Section 4.1) did not lead students to do the retrieval task differently and in consequence did not enhance learning. An alternative explanation is that there was a trade-off between positive effects of gamified feedback on motivation – which may have caused students to try harder, for example – and negative effects such as distraction, which may have reduced learning. Based on the present data, we cannot distinguish between these explanations. However, we can conclude that adding gamified feedback to retrieval practice did not significantly change recall performance on the delayed posttest.

6.4. Limitations and questions for further research

A number of characteristics and limitations of this study should be discussed. For one, the adaptive retrieval practice system that we used may have created a relatively strong (i.e., motivating and effective) control condition because the difficulty of practice was adapted to user performance, in such a way that each trial was (predicted to be) challenging yet achievable for learners (Sense & van Rijn, 2022; van Rijn et al., 2009). The balance between learner ability and task difficulty is itself a popular topic in literature on game design (e.g., Garris et al., 2002), often in relation to the concept of *flow* (Nakamura & Csikszentmihalyi, 2014): Flow – a rewarding, emotional state that arises when an individual is absorbed in a task (Boyle et al., 2012, p. 772) – is thought to be more likely when the amount of challenge of a task closely matches a person's skills (Keller et al., 2011). Thus, exposing learners to challenging but attainable retrieval tasks, may already have created a rewarding and effective learning task, which perhaps left less room for a positive impact of gamified feedback compared to a learning task that feels more tedious or repetitive (e.g., Mekler et al., 2017). Therefore, effects of gamification might be larger in future studies with non-adaptive retrieval practice.

The points-and-progress condition raised a number of follow-up questions. For instance, we had to choose a certain maximum level of points for the progress bar, taking into account that we did not want to frustrate learners who did not manage to fill the bar but also did not want fast learners to stop practicing early. An interesting question is whether adjustments in the progress bar could motivate learners differently. For instance, we assumed that people aim to fill the whole bar but could have also specified explicitly what constitutes good performance (e.g., 5 steps = good, 6 steps = very good, etc.), adjusted the target score based on participants' prior performance, or allowed

participants to set their own goals. Although we found no prior research making systematic comparisons of such manipulations in the gamification literature, recent research on feedback in learning analytics dashboards could provide a theoretical framework for such follow-up research, as it defines different types of reference frames (e.g., progress: comparing learners current performance to earlier results; social: comparisons to other learners) (e.g., Gallagher et al., 2024). In further exploring effects of gamified feedback, it will also be relevant to measure learner motivation not only through self-reports but to include, for instance, measures of persistence or the willingness to engage in further practice (i.e., actual, voluntary practice time).

For practical applications, an important question is what the impact is of gamified feedback in learners in an educational context and on the long-term. Regarding the context, this study involved adult volunteers whose reasons to participate are unknown but likely involved curiosity about word learning tasks, the hope to win a gift cheque, and the wish to contribute to this study. Adult learners in language classes or students in secondary education have different reasons to engage in vocabulary practice, such as language learning goals or the wish to obtain good grades. These might influence their response to gamified feedback (e.g., positive feedback on progress may be more meaningful). Therefore, replication of this study in an educational context would be desirable, and such individual differences between learners could then be studied as possible moderators of gamification effects.

We also conducted an intervention in a single practice session with one delayed posttest. The delayed posttest is a strength of this study as there is a lack of information about how enduring effects of gamification manipulations are (e.g., the overview study by Sailer and Homner (2020) identified only one study with a delayed posttest). However, game-based learning research indicates that effects of games (compared to non-game control conditions) are larger when players engage in multiple sessions over longer periods of time (Wouters et al., 2013). This might be because initially familiarizing oneself with the learning environment takes time and mental resources away from the core learning task (cf. our findings regarding effects of gamification-first on learning behavior), but pays off when learners continue to play (Wouters et al., 2013). More closely related to the present study, it has been argued that understanding and adapting to a gamified feedback system might take time, such that the full potential of such a system might develop with time (Groening & Binnewies, 2019). Although the feedback used in the present study was simple, it would be relevant to see if its effects on learning outcomes change if applied over a longer period of time. Indeed, beneficial effects of gamification are most relevant in the long run when persistence in practice usually lapses (Bernecker & Ninaus, 2021). Yet, durable effects of gamification are not a given. For instance, Sanchez et al. (2020) found positive effects of adding a progress bar together with other gamification options and encouraging messages to digital quizzes, which university students could take voluntarily to prepare for tests. However, this effect was found only for the first of three quizzes offered, making the authors suggest a possible novelty effect where the motivational influence of gamification declined over time.

As a final note, the gamified feedback in this study was quite discreet. Digital games often provide elaborate feedback in the form of sounds, visuals and animations, which might create a more captivating or more enjoyable experience than what was employed here, and could possibly further increase feelings of competence (Juul 2012; in Mekler, 2016). However, while research has shown positive effects of design choices such as pleasant colors, round shapes, and anthropomorphism of materials on motivation and learning (for an overview of emotional design effects, see Brom et al., 2018), the effects of esthetics of game elements on learning are still poorly understood (Javora et al., 2019). Tentatively, more prominent feedback in this study might enhance both the positive effects on task motivation and the potential for distracting users from the core learning task.

6.5. Conclusion

Retrieval practice is an effective learning strategy to promote the long-term retention of factual information (Agarwal et al., 2021; Karpicke, 2017; Rowland, 2014), in particular when practice is repeated over time and adjusted to a learner's level of knowledge (van der Velde, Sense, Borst, & van Rijn, 2021). In the present study, gamified feedback enhanced learners' self-reported motivation for this effective but effortful learning task, both when a simple score was used that increased for correct answers, and when the score was combined with a colorful visualization on a progress bar. Findings for learning behaviors suggest that it is important to consider the potential impact of gamified feedback on learners' experienced load when presenting it in unfamiliar tasks, but there were no effects on learning outcomes on the posttest. Overall, gamified feedback is a potentially useful addition to make retrieval practice more appealing.

CRediT authorship contribution statement

Gesa S. E. van den Broek: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Stefanie Scholten:** Project administration, Methodology, Investigation. **Bente van Thuil:** Writing – original draft, Project administration, Methodology, Investigation. **Hedderik van Rijn:** Writing – review & editing, Supervision, Formal analysis. **Tamara van Gog:** Writing – review & editing, Supervision, Conceptualization. **Maarten van der Velde:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships:

We used an algorithm for spaced repetition in our experiment that is included in similar form in the educational software MemoryLab. Hedderik van Rijn reports a relationship with MemoryLab that includes: board membership, employment, and equity or stocks. Maarten van der Velde reports a relationship with MemoryLab that includes: employment and equity or stocks. The affiliations of HR and MV with the company that publishes MemoryLab has not influenced the setup, analysis, or reporting of the study. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank Dana Hogeveen, Luka van den Herik, and Karlijn Born for their help in constructing the materials and conducting the experiment.

Appendix I. Drop-out and Exclusion Criteria

This section summarizes the dropout over the course of practice. Of the 203 times that users provided consent to start the experiment, 21 times, the entry was subsequently abandoned before starting the retrieval task; 10 entries were removed because participants had completed only one round of retrieval practice or because they started the experiment multiple times. One participant was removed because s/he completed only 30 trials (which is less than M-3*sd = 64.4 trials); five participants were excluded because they reported cheating during practice (by writing information down or looking up information online). The remaining 166 entries for Session 1 were analyzed, although due to technical problems, the survey data of four participants was not complete. Moreover, 14 participants did not return for Session 2, leaving 152 entries with posttest data.

# before	Description	# after
203 started the experiment	4 people abandoned before entering ID 17 abandoned before start of retrieval practice	182
182	7 people did the session too many/few times (5* +2 + 3 = 10 entries)	172
172	1 person completed too few practice trials.	171
171	5 participants admitted cheating	166 in Practice data (survey data incomplete for four)
166	14 people did not do the test	152 in Test data

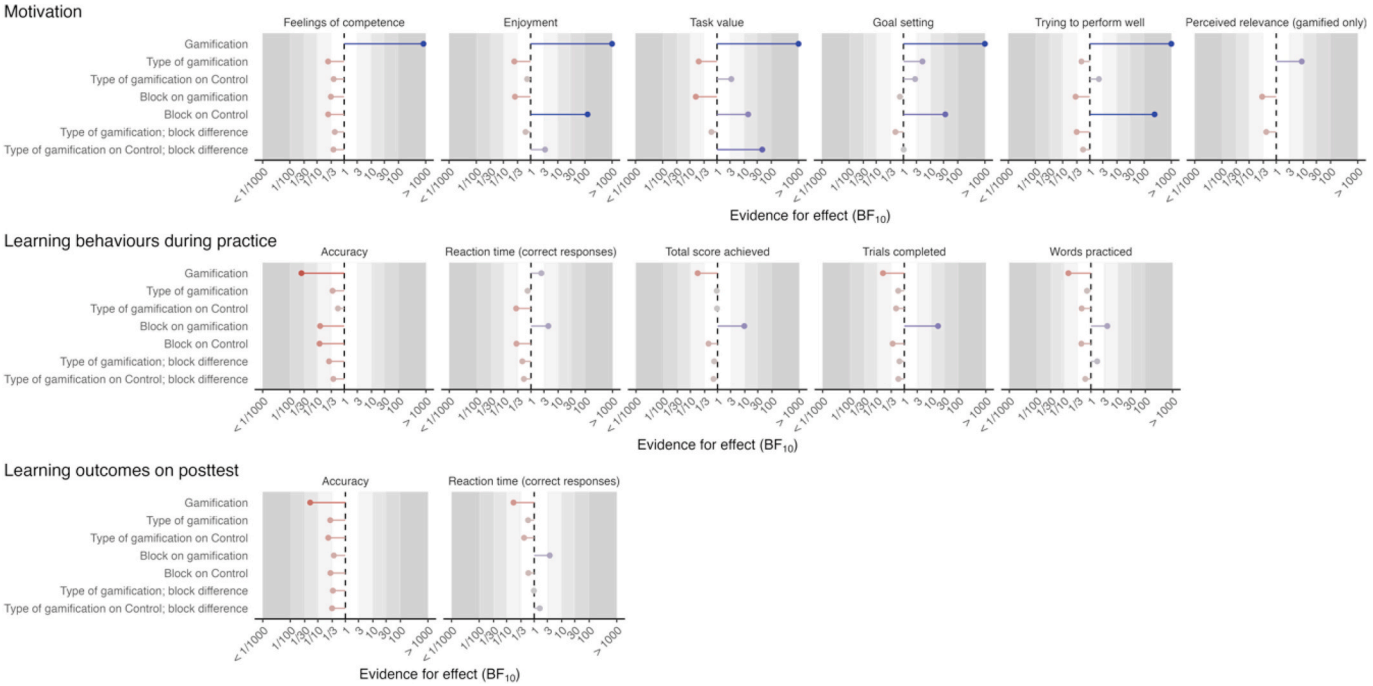
* Four learners dropped out during the control condition and one learner dropped out during the points + progress condition.

Appendix II. Predictors Used in Regression Models

Regression coefficient	Interpretation of effect
Intercept	The model's intercept estimates the average outcome in the Control condition, across both experimental groups and both block orders.
GamifiedTRUE	Effect of Gamification (Gamified - Control)
gamifiedTRUE:exp_group_c	Effect of Type of gamification (Point-and-progress - Points)
gamifiedFALSE:exp_group_c	Effect of Type of gamification on Control (Point-and-progress - Points)
gamifiedTRUE:gamified_first_c	Effect of Block on Gamified Condition (Block 1 - Block 2)
gamifiedFALSE:gamified_first_c	Effect of Block on Control Condition (Block 2 - Block 1)
gamifiedFALSE:exp_group_c: gamified_first_c	Effect of Type of gamification on Control; block difference (Block 2 - Block 1)
gamifiedTRUE:exp_group_c:gamified_first_c	Effect of Type of gamification; block difference (Block 2 - Block 1)

Appendix III. Visual Summary of Evidence Strength across Outcomes

Evidence per outcome for each regression model coefficient being different from zero.



Note: Plot displays evidence in favour of an effect (BF_{10}) for each coefficient (β) of the Bayesian regression models reported in the text. Each subplot corresponds to an outcome variable. Labels along the vertical axis describe coefficients; values along the horizontal axis denote the strength of ev-

idence on a logarithmic scale. Values of $BF_{10} > 1$ (in blue) indicate that the data are more likely under the alternative hypothesis ($\beta \neq 0$); values of $BF_{10} < 1$ (in red) indicate that the data are more likely under the null hypothesis ($\beta = 0$). Longer lines represent stronger evidence.

Data availability

The manuscript contains a link to a repository containing the data and analysis code.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review, 33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arieli-Attali, M. (2016). Self adapted testing as formative assessment: Effects of feedback and scoring on engagement and performance. https://search.proquest.com/docview/1766154660.

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review, 30*, Article 100322. https://doi.org/10.1016/j.edurev.2020.100322

Bernecker, K., & Ninaus, M. (2021). No pain, no gain? Investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior, 114*, Article 106542. https://doi.org/10.1016/j.chb.2020.106542

Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at schmidt and bjork (1992). *Perspectives on Psychological Science, 13*(2), 146–148. https://doi.org/10.1177/1745691617690642

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior, 28*(3), 771–780. https://doi.org/10.1016/j.chb.2011.11.020

Brom, C., Stárková, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review, 25*, 100–119. https://doi.org/10.1016/j.edurev.2018.09.004

Bürkner, P.-C. (2017). Brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software, 80*(1). https://doi.org/10.18637/jss.v080.i01

Carpenter, S. K., Endres, T., & Hui, L. (2020). Students' use of retrieval in self-regulated learning: Implications for monitoring and regulating effortful learning experiences. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09562-w

Center for Selfdetermination Theory. (n.d.). Intrinsic Motivation inventory (IMI). https://selfdeterminationtheory.org/intrinsic-motivation-inventory/.

Chan, E., Nah, F. F.-H., Liu, Q., & Lu, Z. (2018). Effect of gamification on intrinsic motivation. In F. F.-H. Nah, & B. S. Xiao (Eds.), *HCI in business, government, and organizations* (pp. 445–454). Springer International Publishing. https://doi.org/10.1007/978-3-319-91716-0_35.

Chevalier, N. (2018). Willing to think hard? The subjective value of cognitive effort in children. *Child Development, 89*(4), 1283–1295. https://doi.org/10.1111/cdev.12805

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review, 32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. *Human-Computer Interaction, 30*(3–4), 294–335. https://doi.org/10.1080/07370024.2014.993471

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education, 14*(1), 9. https://doi.org/10.1186/s41239-017-0042-5

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

Gallagher, T., Slof, B., van der Schaaf, M., Toyoda, R., Tehreem, Y., Garcia Fracaro, S., & Kester, L. (2024). Reference frames for learning analytics dashboards: The progress and social reference frame and occupational self-efficacy. *Journal of Computer Assisted Learning, 40*(2), 742–760. https://doi.org/10.1111/jcal.12912

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116*(2), 439–453. https://doi.org/10.1037/a0015251

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

Groening, C., & Binnewies, C. (2019). "Achievement unlocked!"—The impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior, 97*, 151–166. https://doi.org/10.1016/j.chb.2019.02.026

Higashi, R., & Schunn, C. D. (2020). Perceived relevance of digital badges predicts longitudinal change in program engagement. *Journal of Educational Psychology, 112*(5), 1020–1041. https://doi.org/10.1037/edu0000401

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology, 28*(4), 597–606. https://doi.org/10.1002/acp.3032

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7*(5), 450–463. https://doi.org/10.1177/1745691612454134

Javora, O., Hannemann, T., Stárková, T., Volná, K., & Brom, C. (2019). Children like it more but don't learn more: Effects of esthetic visual design in educational games. *British Journal of Educational Technology, 50*(4), 1942–1960. https://doi.org/10.1111/bjet.12701

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory* (Vol. 2, pp. 487–514). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kelle, S., Klemke, R., & Specht, M. (2013). Effects of game design patterns on basic life support training content. *Journal of Educational Technology & Society, 16*(1), 275–285.

Keller, J., Ringelhan, S., & Blomann, F. (2011). Does skills–demands compatibility result in intrinsic motivation? Experimental test of a basic notion proposed in the theory of flow-experiences. *The Journal of Positive Psychology, 6*(5), 408–417. https://doi.org/10.1080/17439760.2011.604041

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology, 115*, Article 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management, 45*, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Krath, J., Schürmann, L., & von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior, 125*, Article 106963. https://doi.org/10.1016/j.chb.2021.106963

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Kwan, H. O. K., & Lee, C. (2013). Re-Examining the goal-setting questionnaire. In *New developments in goal setting and task performance* (pp. 607–624). Routledge. https://doi.org/10.4324/9780203082744-53.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752–768. https://doi.org/10.1177/1046878114563660

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior, 71*, 508–515. https://doi.org/10.1016/j.chb.2015.08.008

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

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

Mayer, R. E. (2020a). Cognitive foundations of Game– based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 83–110). MIT Press. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=6018189.

Mayer, R. E. (2020b). *Multimedia Learning. Higher education from.* Cambridge University Press; Cambridge University Press. https://doi.org/10.1017/9781316941355

Mekler, E. D. (2016). The motivational potential of digital games and gamification—The relation between game elements. *Experience and Behavior Change*. https://doi.org/10.5451/UNIBAS-006625130

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). 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

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition, 37*(8), 1077–1087. https://doi.org/10.3758/MC.37.8.1077

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). BayesFactor: Computation of bayes factors for common designs [Computer software]

Version 0.9.12-4.7. https://cran.r-project.org/web/packages/BayesFactor/index.html.
Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology: The collected works of Mihaly csikszentmihalyi* (pp. 239–263). Netherlands: Springer. https://doi.org/10.1007/978-94-017-9088-8_16.
Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2016). The higher the score, the higher the learning outcome? Heterogeneous impacts of leaderboards and choice within educational videogames. *Computers in Human Behavior, 65*, 391–401. https://doi.org/10.1016/j.chb.2016.08.042
Nebel, S., Schneider, S., Beege, M., & Rey, G. D. (2017). Leaderboards within educational videogames: The impact of difficulty, effort and gameplay. *Computers & Education, 113*, 28–41. https://doi.org/10.1016/j.compedu.2017.05.011
Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*(2), 101–117. https://doi.org/10.1037/1076-898X.14.2.101
Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria [Computer software] http://www.R-project.org/.
Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003
Rivers, M. L. (2020). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09578-2
Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003
Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x
Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020
Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w
Sanchez, D. R., Langer, M., & Kaur, R. (2020). Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education, 144*, Article 103666. https://doi.org/10.1016/j.compedu.2019.103666
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
Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science, 8*(1), 305–321. https://doi.org/10.1111/tops.12183
Sense, F., & van Rijn, H. (2022). Optimizing fact-learning with a response-latency-based adaptive system. *PsyArXiv*. https://doi.org/10.31234/osf.io/chpgv
Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review, 32*(3), 707–734. https://doi.org/10.1007/s10648-020-09522-4
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer. https://doi.org/10.1007/978-1-4419-8126-4
Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
Tondello, G., Premsukh, H., & Nacke, L. (2018). A theory of gamification principles through goal-setting theory. *Hawaii international conference on system sciences*. https://doi.org/10.24251/HICSS.2018.140
van den Broek, G. S. E., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education, 5*(2), 52–66. https://doi.org/10.1016/j.tine.2016.05.001
van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *The R Journal, 6*(1), 111–122.
van der Velde, M., Sense, F., Borst, J., & van Rijn, H. (2021). Alleviating the cold start problem in adaptive learning using data-driven difficulty estimates. *Computational Brain and Behavior, 4*(2), 231–249. https://doi.org/10.1007/s42113-021-00101-6
van der Velde, M., Sense, F., Spijkers, R., Meeter, M., & van Rijn, H. (2021). Lockdown learning: Changes in online foreign-language study activity and performance of dutch secondary school students during the covid-19 pandemic. *Frontiers in Education, 6*. https://www.frontiersin.org/articles/10.3389/feduc.2021.712987.
van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In A. Howes, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 9th international conference on cognitive modeling* (pp. 110–115).
Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive Psychology, 60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001
Weissgerber, S. C., & Rummer, R. (2023). More accurate than assumed: Learners' metacognitive beliefs about the effectiveness of retrieval practice. *Learning and Instruction, 83*, Article 101679. https://doi.org/10.1016/j.learninstruc.2022.101679
Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*(2), 217–266. https://doi.org/10.1177/1046878108321866
Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. https://doi.org/10.1037/a0031311
Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review, 30*, Article 100326. https://doi.org/10.1016/j.edurev.2020.100326