Nudging Adolescents Towards Recommended Maths Exercises With Gameful Rewards

Jeroen Ooge*[1], Joran De Braekeleer*[2], and Katrien Verbert[3]

[1] Utrecht University, Department of Information and Computing Sciences, Utrecht, the Netherlands.
j.ooge@uu.nl
[2] KU Leuven, Department of Computer Science, Leuven, Belgium.
joran.debraekeleer@gmail.com
[3] KU Leuven, Department of Computer Science, Leuven, Belgium.
katrien.verbert@kuleuven.be

Abstract. E-learning systems that force learners to solve personalised exercises lower their control and possibly their motivation. To better balance learner control and automation, we created an app for practising high school equation-solving in which learners can select exercises from tailored sets and are nudged towards recommended ones with gameful rewards. Furthermore, labels indicate exercises' difficulty. A randomised controlled experiment with 154 adolescents revealed that our nudges made learners select harder exercises without negatively impacting short-term learning performance and self-reported competence. However, difficulty labels did not have such effects. In sum, our study suggests that reward-based nudging is promising to let learners voluntarily engage in more challenging learning materials while preserving selection freedom.

Keywords: Learner control · Nudging · Gamification.

1 Introduction

Adaptive e-learning systems try to improve learning processes by automatically tailoring learning materials to learners' mastery levels. To challenge learners without overwhelming or boring them, exercises of suitable difficulty should keep learners within the so-called *zone of proximal development* [11]. As a result, e-learning platforms often enforce exercises that are deemed most effective by an algorithm. However, this approach can lower learners' intrinsic motivation due to reduced freedom of choice as autonomy is a core pillar in self-determination theory [5]. Alternatively, learners can be presented with multiple recommended exercises and *nudged* [2,8] towards exercises in the zone of proximal development without taking away their freedom of choice. Yet, a recent review showed that

* Both authors contributed equally to this research. The Version of Record of this contribution is published in Artificial Intelligence in Education 2024, Lecture Notes in Computer Science (vol 14830 pp. 328–335), and is available online at https://doi.org/10.1007/978-3-031-64299-9_28

nudging has rarely been studied in the context of recommender systems [7]. Moreover, there is untapped potential in *smart nudges*, which tailor nudges to individuals and their context [7,8]. This raises a first research question:

RQ1. How can smart nudging be operationalised in an educational recommender system to nudge learners towards recommended exercises?

Many types of nudges can leverage cognitive biases or social norms to steer people's behaviour in a desirable direction [6]. For example, incentive nudges take advantage of people's loss aversion and have been operationalised with gameful rewards to increase persuasiveness [14], which hints at a link between nudging and *gamification* [1,15]. Furthermore, salience nudges focus people's attention on what seems relevant to them. Difficulty labels for learning materials, for example, can discourage learners from processing materials indicated as difficult unless they are given task-related choice beforehand [16]. Overall, previous studies on nudging in education have shown mixed effects: while nudging often successfully changes learners' behaviour, not every nudge is equally effective in all contexts [2]. This leads to our second research question:

RQ2. How do nudges in the form of gameful rewards and difficulty labels affect chosen exercise difficulty, learning performance, and self-perceived competence?

Our work contributes to answering the above research questions. Specifically, we designed and developed a smartphone app for high school students to practice equation-solving, incorporating automated recommendations based on skill-level Elo ratings and smart gameful rewards that nudge learners towards exercises whose difficulty lies in the zone of proximal development. A randomised controlled experiment with 154 adolescents shows that gameful rewards can nudge learners towards harder exercises without decreasing their short-term learning performance or self-reported competence. Overall, we hope our work sparks more interest in personalised learning systems that motivate adolescents.

2 Materials and Methods

This section briefly introduces our app and overall study procedure. Our study was approved by the ethical committee of KU Leuven (reference G-2023-6197).

2.1 Smartphone Application

Figure 1 shows our app's general workflow. Upon choosing a topic, the app generates four exercises with varying difficulties using the algorithm described below and presents them in random order on a selection screen. There, learners first select one of four reward types, which they collect upon solving an exercise: stars to climb up in a leaderboard, coins to unlock badges, chests to collect objects, and fish to feed a virtual cat. These reward types correspond to game-design elements that are in theory most preferred by the four most common

Hexad gamification user types [1]. Next, learners choose an exercise and get three chances to solve it: they type the answer or an intermediate step and get direct feedback on whether their input is correct. Learners who find the answer are rewarded and can visit the corresponding reward interface or continue practising. This workflow results from an iterative design process with 12 young adults, detailed in [3].

Generating Exercises. Solving linear, quadratic, and cubic equations requires computational skills such as addition, multiplication, changing signs, distribution, and applying the discriminant or Horner's method. Our app captures these skills in 20 templates (e.g., $ax^2+b=c$) and generates exercises by randomly picking parameters and solutions. Exercises with difficulties slightly above learners' mastery level yet within the zone of proximal development are recommended. To estimate how difficult exercises are for specific learners, the app uses an Elo rating system [13] inspired by the variant typical for chess. Concretely, all *skills* have a global Elo rating, and learners have personal ratings for each skill, which gives a fine view of how well they master equation-solving *skills*. The rating of exercises is defined as the highest rating of the skills in their template.

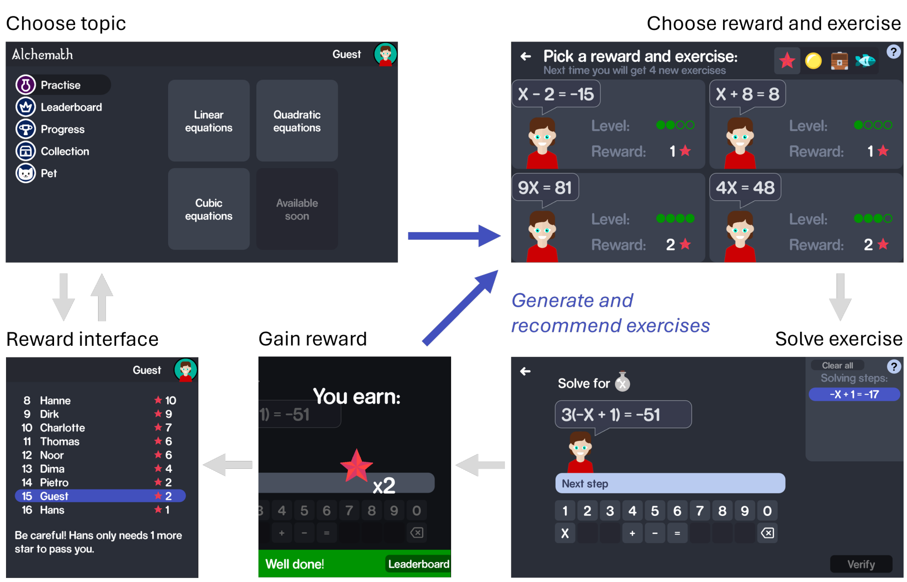

Fig. 1. Workflow in our app: learners choose a topic, reward, and exercise; after solving the latter, they are rewarded and visit the reward interface or continue practising.

2.2 Participants and Study Procedure

We asked high school maths teachers in Belgium (Flanders) to let students participate in our study during class without pressuring them and while providing exercises on paper for students who did not participate. Interested students provided informed consent, and those below 16 needed parental consent. Our study was a randomised controlled experiment with the four groups in Figure 2, using a 2×2-design with the following variables:

– **Difficulty labels**: groups D1I0 and D1I1 saw exercises accompanied by coloured dots indicating their difficulty; the other groups did not.
– **Increased rewards**: groups D0I1 and D1I1 gained two rewards for recommended exercises; the other groups received one reward for all exercises.

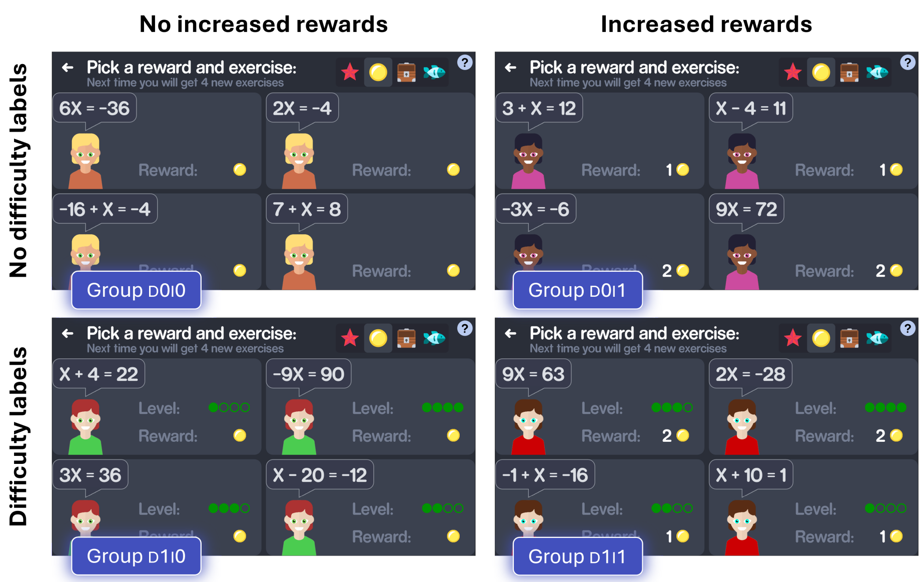

Fig. 2. The differences between selection screens based on increased rewards and difficulty labels that differentiated the four test groups.

Participants could freely practise topics. After 15 exercises, they were referred to the post-study questionnaire in Table 1: questions Q1–Q5 measured *perceived competence* with a subscale of the Intrinsic Motivation Inventory [10] and question Q6 asked what participants looked at when choosing exercises. Afterwards, participants could continue using the app. In the background, we logged Elo rating changes and details of the exercises participants selected.

Table 1. Post-study questionnaire where Q1–Q5 were scored on a 7-point scale.

ID	Question
Q1	I think I am pretty good at this task
Q2	I think I did pretty well at this activity, compared to other students
Q3	I am satisfied with my performance at this task
Q4	I felt pretty skilled at this task
Q5	After working at this task for a while, I felt pretty competent
Q6	When I chose an exercise, I particularly looked at: the reward / the exercise / the level / other; explain [in an open text field]

3 Results

In total, 154 adolescents participated in the study. To obtain a more focused analysis, we limited our sample to 127 adolescents between 13 and 19 years old who answered at least 10 exercises: 28 ended up in D0I0, 34 in D0I1, 29 in D1I0, and 36 in D1I1. Participants identified as female (61%), male (32%), or different (6%). During the experiment, the average participant solved 67 exercises (SD = 93) in roughly one hour, primarily practising linear equations. Only 84 participants filled out the post-study questionnaire.

Chosen Exercises. Figure 3 shows that participants with increased rewards chose harder exercises on average. A two-way ANOVA confirmed that having increased rewards significantly affected chosen difficulty ($p < 0.01$) in contrast to seeing difficulty labels ($p = 0.91$). Moreover, although borderline, there was no interaction effect between seeing increased rewards and difficulty labels ($p = 0.05$). To further test one-sided differences, we conducted pairwise t-tests with Benjamini-Hochberg correction, which confirmed that D0I1 and D1I1 chose harder exercises than D0I0 (both $p \leq 0.01$); other comparisons were insignificant (all $p > 0.12$). Finally, the answers to Q6 revealed that participants often considered rewards while selecting exercises, especially in groups with increased rewards: 75% and 70% of the participants in D0I1 and D1I1 indicated to pay attention to the rewards, respectively, compared to 36% and 47% in D0I0 and D1I0, respectively. Participants appreciated rewards because they *"motivated [them]"*, *"were fun"*, or allowed them to progress on the gameful interfaces linked to the reward types.

Short-Term Learning Performance. We investigated learning performance in two ways. First, we measured the overall change in Elo rating during the experiment. While groups with increased rewards had slightly higher Elo gains (see Figure 3), a two-way ANOVA showed these differences were insignificant (all $p > 0.30$). Second, we defined learning performance in terms of the correctness of the given answers. For each finished exercise, we assigned a performance score of 0 if participants gave three wrong answers or gave up and 1/# attempts otherwise. The average performance score in all groups was about 0.88, meaning exercises were often solved on the first attempt. A two-way ANOVA did not find significant differences across groups (all $p > 0.13$).

Impact on Self-Reported Competence. Participants in groups with and without increased rewards scored their competence for solving equations in our app with a 4.75 and 5 out of 7 on average, respectively. This difference was, however, insignificant according to a two-way ANOVA (all $p > 0.24$).

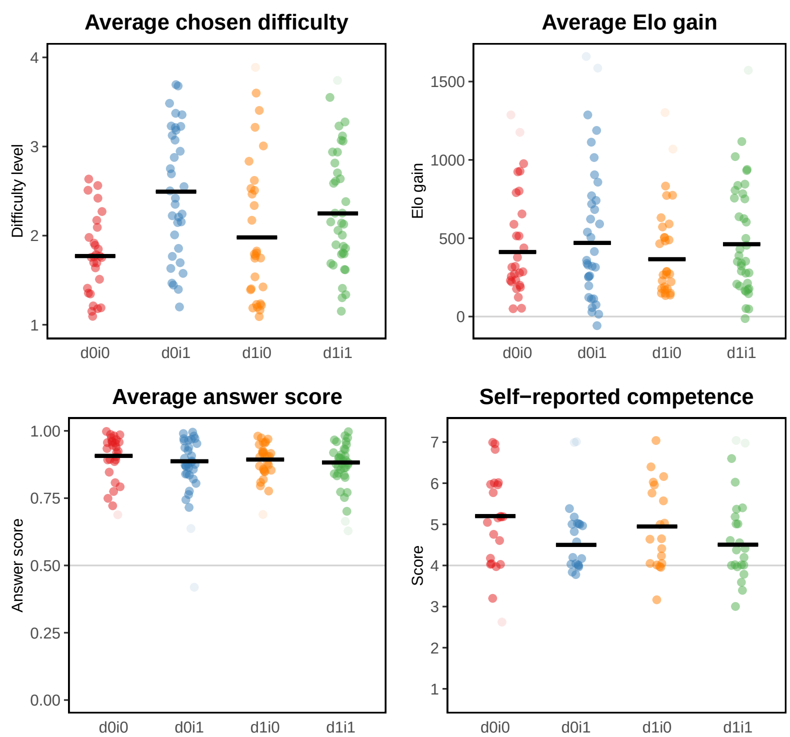

Fig. 3. Scatter plots of the three studied metrics: average chosen difficulty, learning performance measured with average Elo gain and average answer score, and self-reported competence. Group outliers are faded, and horizontal bars indicate group means.

4 Discussion and Conclusions

We explored the space between fully automated e-learning systems and systems wherein learners have full control since neither seems optimal: full automation might reduce intrinsic motivation for learning due to reduced autonomy [5], whereas full control over which exercises to solve is problematic if learners systematically under- or overestimate themselves. Furthermore, shared control can yield better learning outcomes [9] and increase learners' trust in recommender systems [12]. Our intermediate solution uses gameful rewards as nudges towards recommended exercises and displays difficulty labels to support decision-making.

4.1 Smart Nudging With Gameful Rewards Is Feasible

Few nudging interventions positively affect everyone [2]. In our case of gameful reward-based nudging in an educational recommender, the power of automatically recommending learning materials can diminish if learners are not persuaded by the rewards. As an initial step towards avoiding this pitfall, we operationalised *smart* nudging [8] by letting learners select their preferred reward type (see RQ1). Our results suggest that this nudging indeed persuades adolescents to pick more challenging exercises without negatively affecting their short-term learning performance and self-assessed competence (see RQ2). Furthermore, we found that labels indicating exercises' estimated difficulty did not yield such effects.

Future work could combine our ideas with research on personalising gamification [15] to automatically deduce learners' preferred rewards and study whether this enhances nudging effects and desirable learning goals such as performance and motivation. In addition, the trade-offs of reward-based nudging should be further explored. On the one hand, gameful rewards can be an example of transparent nudging, which is relevant to the ethical debate around nudging in the sensitive context of education for adolescents [8]. On the other hand, the motivational aspects of gameful rewards should be studied in more detail as they might mainly tap into extrinsic motivation, which has been criticised for potentially undermining intrinsic motivation [4].

4.2 Limitations and Future Work

While our findings are promising, our study had several limitations. First, most participants practised linear equations, which were relatively easy, as evidenced by the overall high performance scores. Longer-term experiments with harder exercises should verify whether our findings hold. Additionally, although the effect was not statistically significant, we are mindful that groups with increased rewards reported lower competence while all groups performed equally well. Future studies could investigate whether this undesirable phenomenon occurs in larger samples. Finally, the large importance that participants dedicated to rewards when selecting exercises might have been reinforced by the classroom context. For example, participants often whipped each other up for the leading position on the leaderboard. Future studies could study how such a heated atmosphere impacts learning outcomes and learners who feel less motivated by rewards popular among their peers.

In sum, we hope follow-up studies further explore how e-learning systems best balance automation and learner control to foster motivation and learning.

References

1. Altmeyer, M., Tondello, G.F., Krüger, A., Nacke, L.E.: HexArcade: Predicting Hexad User Types By Using Gameful Applications. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. pp. 219–230. CHI PLAY

'20, Association for Computing Machinery, New York, NY, USA (Nov 2020). https://doi.org/10.1145/3410404.3414232
2. Damgaard, M.T., Nielsen, H.S.: Nudging in education. Economics of Education Review **64**, 313–342 (Jun 2018). https://doi.org/10.1016/j.econedurev.2018.03.008
3. De Braekeleer, J.: Gebruikers nudgen naar effectieve oefeningen. Master's thesis, KU Leuven, Faculty of Engineering Science, Leuven, Belgium (2023)
4. Deci, E.L., Koestner, R., Ryan, R.M.: Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. Review of educational research **71**(1), 1–27 (2001)
5. Deci, E.L., Ryan, R.M.: Motivation, personality, and development within embedded social contexts: An overview of self-determination theory. The Oxford handbook of human motivation **18**(6), 85–107 (2012)
6. Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., Vlaev, I.: Influencing behaviour: The mindspace way. Journal of Economic Psychology **33**(1), 264–277 (Feb 2012). https://doi.org/10.1016/j.joep.2011.10.009
7. Jesse, M., Jannach, D.: Digital nudging with recommender systems: Survey and future directions. Computers in Human Behavior Reports **3**, 100052 (Jan 2021). https://doi.org/10.1016/j.chbr.2020.100052
8. Karlsen, R., Andersen, A.: Recommendations with a Nudge. Technologies **7**(2), 45 (Jun 2019). https://doi.org/10.3390/technologies7020045
9. Long, Y., Aleven, V.: Mastery-Oriented Shared Student/System Control Over Problem Selection in a Linear Equation Tutor. In: Micarelli, A., Stamper, J., Panourgia, K. (eds.) Intelligent Tutoring Systems. pp. 90–100. Lecture Notes in Computer Science, Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-39583-8$_9$
10. McAuley, E., Duncan, T., Tammen, V.V.: Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. Research Quarterly for Exercise and Sport **60**(1), 48–58 (Mar 1989). https://doi.org/10.1080/02701367.1989.10607413
11. Murray, T., Arroyo, I.: Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. In: Cerri, S.A., Gouardères, G., Paraguaçu, F. (eds.) Intelligent Tutoring Systems. pp. 749–758. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2002). https://doi.org/10.1007/3-540-47987-2$_7$5
12. Ooge, J., Dereu, L., Verbert, K.: Steering Recommendations and Visualising Its Impact: Effects on Adolescents' Trust in E-Learning Platforms. In: Proceedings of the 28th International Conference on Intelligent User Interfaces. pp. 156–170. IUI '23, Association for Computing Machinery, New York, NY, USA (Mar 2023). https://doi.org/10.1145/3581641.3584046
13. Pelánek, R.: Applications of the Elo rating system in adaptive educational systems. Computers & Education **98**, 169–179 (Jul 2016). https://doi.org/10.1016/j.compedu.2016.03.017
14. Petrykina, Y., Schwartz-Chassidim, H., Toch, E.: Nudging users towards online safety using gamified environments. Computers & Security **108**, 102270 (Sep 2021). https://doi.org/10.1016/j.cose.2021.102270
15. Rodrigues, L., Toda, A.M., Oliveira, W., Palomino, P.T., Vassileva, J., Isotani, S.: Automating Gamification Personalization to the User and Beyond. IEEE Transactions on Learning Technologies **15**(2), 199–212 (Apr 2022). https://doi.org/10.1109/TLT.2022.3162409

16. Schneider, S., Nebel, S., Meyer, S., Rey, G.D.: The interdependency of perceived task difficulty and the choice effect when learning with multimedia materials. Journal of Educational Psychology **114**(3), 443–461 (2022). https://doi.org/10.1037/edu0000686