

VU Research Portal

Student engagement with computerized practising

Cornelisz, Ilja; van Klaveren, Chris

published in

Journal of Computer Assisted Learning
2018

DOI (link to publisher)

[10.1111/jcal.12292](https://doi.org/10.1111/jcal.12292)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Cornelisz, I., & van Klaveren, C. (2018). Student engagement with computerized practising: Ability, task value, and difficulty perceptions. *Journal of Computer Assisted Learning*, 34(6), 828-842.
<https://doi.org/10.1111/jcal.12292>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Student engagement with computerized practising: Ability, task value, and difficulty perceptions

Ilja Cornelisz  | Chris van Klaveren

Faculty of Behavioural and Movement Sciences - Department of Education Science - Amsterdam Center for Learning Analytics, Vrije Universiteit Amsterdam, The Netherlands

Correspondence

Ilja Cornelisz, Behavioural and Movement sciences - section Educational Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

Email: i.cornelisz@vu.nl

Funding information

The Netherlands Organisation for Scientific Research (NWO), Grant/Award Number: 405-14-510

JEL Classification: A20; I21; L86

Abstract

A prerequisite for low-stakes activities to improve learning is to keep students engaged when confronted with challenging material. Comparing a personalized and non-personalized version of computerized practising, this study experimentally evaluates the relationships between student effort and ability across different dimensions of task perceptions. Students practise longer when the task is perceived to be not too difficult. Students assigned to a personalized version of the tool have a lower success rate while practising, but this does not translate to differences in practice intensity, task perceptions, or summative test scores. In a personalized practising environment, perceived interest and usefulness both have the potential to promote engagement, albeit in different ways. Students with a lower level of subject-specific ability find the tool more interesting and, particularly, consider the non-personalized tool difficult but useful. Instead, in the personalized condition, it is the group of students with higher prescores who value the tool as relatively useful. These results indicate that more students will remain engaged with adaptive practising if software takes into account such differences. Multiple approaches and algorithms may thus be necessary to optimally adapt practising to individual learners.

KEYWORDS

computerized practising, personalization, task difficulty, task value

1 | INTRODUCTION

Educators are generally optimistic that adaptive software in education has the potential to boost student learning outcomes (Demske, 2012). Currently, however, education-related software products tend to have mixed effects in terms of benefiting student learning (Bulman & Fairlie, 2016; Slavin, 2002, 2004). A recent overview points to a variety of reasons to help explain the current ineffectiveness of computerized practising, such as a lack of necessary digital skills of both teachers and students, naive policy implementation strategies, poor understanding of the underlying pedagogy, or simply

the poor quality of the educational software used (OECD, 2015). In listing the various ways in which advancements can be made, the report concludes:

When digital tools support students' engagement with challenging material, thus extending learning time and practice, or help students to assume control over the learning situation, by individualising the pace with which new material is introduced or by providing immediate feedback, students probably learn more. (OECD, 2015, p. 166)

The authors declare that this research is part of a project (nr. 405-14-510) funded by The Netherlands Organisation for Scientific Research (see: <http://www.nwo.nl/en/about-nwo/what+does+nwo+do/funding>).

This highlights that student engagement is an important prerequisite for learning to occur, or as Newmann (1992) puts it, "Students cannot be expected to achieve unless they concentrate, work, and invest themselves in the mastery of school tasks" (p.3). A better understanding of how student ability and perceptions regarding task difficulty and task value interact is vital to improve student engagement and performance in differentiated educational settings (Tomlinson et al., 2003). Educators have long argued that modern information and communication technologies can facilitate reforms towards a more effective and individualized learning environment (Demski, 2012; Lou, Abrami, & d'Apollonia, 2001), for example, through adaptive learning activities (Kulik, 2003). Recent reviews have evaluated this potential; and the findings indicate, among other things, that mobile applications can only improve learning strategies and outcomes when implemented with a rigid educational underpinning (Jeng et al., 2010), that tablets particularly promote learning outcomes if usability is high (Hassler, Major, & Hennessy, 2016), and that individual learning styles importantly affect how adaptive digital learning environments are to be designed (Truong, 2016). Importantly, the current evidence is ambiguous and more experimental research is required to improve the knowledge base regarding whether and how computers can help engage schoolchildren in learning (Bulman & Fairlie, 2016) and how modern technologies can be exploited to enhance learners' motivation (Ciampa, 2014).

In this study, we aim to find an answer to how task perceptions regarding low-stakes computerized practising relate to observed patterns of student engagement and performance for students of different ability. Furthermore, we evaluate the impact of personalization on task perceptions and how differences in perceptions and effort evolve over time. Both aspects (i.e., interlinkages between task perceptions and learner readiness and effects of personalization across a heterogeneous group of learners) are unresolved issues in the empirical literature (Tomlinson et al., 2003; Truong, 2016).

This study contributes to this literature by evaluating the relationships between the effort exerted in computerized practising, student ability, and perceptions related to task value and task difficulty constructs based on a large field experiment. The results provide causal evidence for whether students exert more effort in computerized practising when the process itself is personalized, and whether it impacts task perceptions differentially for different groups of students. Lastly, this study therefore also examines whether increases in the level of effort positively affect performance on summative tests.

This research focuses on 455 secondary school students collected throughout the school year 2014–2015. During this period, the academic performance and computerized practising intensity of students was monitored for four different subjects (Dutch, Biology, Economics, and History). Afterwards, students were asked about their experiences with computerized practising, making it possible to link observed differences in practice intensity and student ability to questionnaire items related to constructs of task difficulty and task value.

Both a personalized and non-personalized version of computerized practising are evaluated in this study. When students practice with the personalized practising programme relative practising performance, knowledge type, difficulty level, and mastery learning are taken into account. When the practising process is not personalized, a

predetermined sequence of exercises is offered, which, at least in theory, is the representative for the upcoming summative test.

1.1 | Practising and student engagement

Despite some obvious challenges in keeping school-aged adolescents engaged throughout the learning process, there are some clear indications regarding phenomenological, instructional, teacher, individual, and school factors on how to avoid potential alienation and disconnection (Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2014). Relevant for the process of practising content is that student engagement notably differs with respect to whether they like the task at hand or they consider it to be challenging and important. Following the principles of flow theory, students should simultaneously experience concentration, interest, and enjoyment when practising in order for a more complex set of subject-specific skills to develop (Csikszentmihalyi, 1997).

Given that computerized content practising is often implemented as a low-stakes activity, students should consider the task worth doing for its own sake. Expectancy-Value Theory, first developed by Atkinson (1957), posits that an individual's choice, persistence, and performance can be explained by their beliefs about how well they will do on a certain activity and the extent to which they value the activity. Whereas some of the task value constructs are in ways similar to other theories of motivation, most notably self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000) they originate from different perspectives and are also distinct in terms of applications. Task value can be described in terms of its three subcomponents interest (i.e., intrinsic value), usefulness (i.e., utility value), and importance (i.e., attainment value), of which the latter two have been found to significantly predict effort and performance on low-stakes tasks (Cole, Bergin, & Whittaker, 2008).

For adolescents working on academic tasks of considerable difficulty, self-esteem maintenance motives are considered pivotal determinants of task value, which, in contrast to predictions for adults, is decreasing in perceived task difficulty (Eccles & Wigfield, 1995). Also, in determining the relative importance of expectancies for success and task value, the latter is considered particularly important in motivating students to engage in low-stakes activities, which is often how computerized practising is implemented in the educational curriculum (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000).

Personalizing the practising process would further promote continuous effort to practise as it better addresses individual learning needs, thereby increasing perceived expectancies for success and task value (Eccles et al., 1983; Tomlinson et al., 2003). On the other hand, personalization can have ambiguous, or even negative, effects on student engagement through, depending on its consequences with respect to task difficulty. Whereas increasing task difficulty generally improves task value for mature learners, this result is not evident for adolescents as they are prone to disengage in settings where tasks are perceived too difficult (Eccles & Wigfield, 1995).

Personalized practising can thus result in lower student engagement, even when the needs of students are better addressed. For instance, a personalized version of computerized practising may return more difficult exercises to students after they have performed

relatively well on previous tasks. Due to this increased difficulty level, students may become frustrated when using the tool and actually become less engaged, even though practising with these more difficult exercises is better for them in improving their overall learning. It is therefore important to empirically establish the intricate relationships between task difficulty, computerized practising, and student performance. Lastly, a recent meta-analysis on educational technology applications in K-12 education indicates that improvements were markedly different for students of low and medium ability (Cheung & Slavin, 2012), suggesting that it is important to estimate results at different levels of (subject specific) student ability.

1.2 | Two approaches towards computerized practising

Table 1 characterizes features of both versions of computerized practising considered in this study. The table shows that both versions recognize that exercises can have different difficulty levels and address different knowledge types. Each exercise that can be offered to the student is therefore labelled with a knowledge type and difficulty level. The knowledge type level indicates the level of learning complexity, based on the widely adopted and revised version of Bloom's taxonomy (Anderson et al., 2001; Bloom et al., 1956; Krathwohl, 2002). The levels considered here are (1) memorization (replication), (2) active use of knowledge (application), and (3) comprehensive understanding (insight).

The difficulty level for each exercise is determined on the success rate of, on average, 3,000 students from previous cohorts that have worked with the tool. Based on this historical success rate, each exercise is assigned to one of three difficulty levels. The first interval represents relatively easy exercises (33.33% highest success rates), the second interval to levels of moderate difficulty (between 33.33% and 66.67% highest success rates), and the third interval are exercises for which the success rate is among the lowest 33.33%. Both versions also provide feedback while practising, which has been found to be one of the most important factors in facilitating learning (Hattie & Timperley, 2007). A student receives feedback immediately after answering an exercise on whether the answer was correct and with cues on where to find information relevant to improve their level of understanding for that particular exercise. Even though each student receives the same type of feedback, it could be that the given feedback to students in the personalized practising version is more effective because it better matches the current knowledge and performance level of the student (Shute, 2008).

TABLE 1 Characterization of both computerized practising conditions

Non-personalized practising	Personalized Practising
Knowledge type	Knowledge type
Difficulty level	Difficulty level
Feedback	Feedback
	Current performance
	Mastery learning
	Effort sustainment

One important difference between both versions is also how difficulty levels are incorporated in the practising process. The non-personalized version distinguishes between these three difficulty interval levels and the aforementioned three knowledge types. Each practising session randomly selects a string of exercises for each topic; orders these exercises subsequently by topic, knowledge type, and difficulty level; and returns this string of exercises in this specific order to the student. It follows that the non-personalized practising condition simulates a more traditional practising approach in which the path of returned exercises is predetermined in terms of topics, knowledge types, and difficulty levels.

The personalized practising condition, however, does not generate a predetermined string of exercises but instead returns exercises contingent on the topic-specific real-time practising performance of the individual student. The practising process is personalized by conditioning on (a) the current individual topic-related understanding level relative to that of the class, (b) mastery learning, and (c) effort sustainment. The individual understanding level reflects the proportion of correct answers on a certain topic. The personalized version considers relative understanding levels because individual understanding levels may reflect the general difficulty level of a particular topic, rather than the poor or below average understanding of the student. The programme updates these relative understanding levels every time a new exercise is assigned. The probability that a particular topic is returned is inversely proportional to these relative understanding levels. In other words, the more difficult a topic is for an individual student, the more likely (s)he will receive an exercise on this topic.

Mastery learning within the context of personalized practising requires that students can only advance to the next knowledge type level when the current level is sufficiently mastered. It has been shown that mastery learning positively influences student achievement (Kulik, Kulik, & Bangert-Drowns, 1990; Slavin, 1987) and positively affects student attitudes towards content and instruction (Kulik et al., 1990). The latter finding suggests that the personalized practising approach has the potential to improve student engagement. The personalized version incorporates this by returning a question of a particular knowledge type before a threshold level of success rate is obtained.

Effort sustainment is an important prerequisite for practising to promote learning. Following *flow theory*, keeping students engaged requires practice material to be just a bit in advance of the student's current level of readiness (Csikszentmihalyi, 1997). The personalized version addresses this by offering exercises of a success rate that is slightly above a student's current level of readiness. If the exercise turns out to be too difficult to complete successfully independently, the assumption is that the task will still be in the zone of proximal development (Vygotsky, 1978) and that the feedback provided by the software provides the student with the necessary scaffolding and support. Studies confirm that students are indeed more likely to sustain efforts when working at tasks at moderate levels of difficulty, compared to when tasks are either too difficult or under challenging (see Tomlinson et al., 2003, for a review). When problems are consistently too difficult, low success rates can refrain students from further practising. For example, empirical research on math anxiety indicate that frustration with practising can be an important mechanism in

explaining poor performance (Ashcraft, Krause, Hopko, Berch, & Mazzocco, 2007). At the same time, when tasks are consistently too easy, students, despite high success rates, are likely to lose motivation and also avoid further practice (National Research Council, 2003).

Practice intensity may thus decrease substantially when exercises offered are either perceived to be too easy (i.e., “boring”) or too difficult (i.e., “frustrating”), which is more likely to occur in the non-personalized version of practising. It is thus relevant to empirically address how practice intensity develops over time and to distinguish between assignment conditions. A detailed technical discussion of the algorithms underlying both computerized practising versions is given in Section 2.3.

2 | METHODS

2.1 | Dutch education system and participants

The general structure of the Dutch educational system is shown in Figure 1. Dutch compulsory schooling law requires children to start with primary education in the year they turn 5 years old. Children are tracked into different secondary education levels upon finishing primary education (Grade 6). This decision is based on the test scores achieved on a standardized nationwide test and on the advice of the primary school teacher. Children can be tracked into three secondary education levels: (1) prevocational education (4 years), (2) general secondary education (5 years), and (3) preuniversity education (6 years). Prevocational education prepares children for vocational education, secondary general education prepares children for universities of applied sciences, and preuniversity education prepares children for universities. This research focuses on students enrolled in preuniversity, upper secondary general education, and prevocational secondary education. All student are enrolled in Grades 7, 8, or 9.

Five secondary schools participated in this experiment with in total 28 different classes. All 654 students were randomly assigned within classes to a personalized and a non-personalized version of the digital practising tool, such that the effects of the personalized practising condition on task perceptions and student performance can be causally evaluated. The practice intensity and performance of these 654 students was tracked from August 2014 to June 2015, in

which over 150,000 exercises were answered in the digital practising environment. Of these students, 618 (94%) filled out the pre-experiment questionnaire regarding background characteristics. Four hundred fifty-five students (70%) also took part in the postexperiment questionnaire, asking them about their experiences with using the tool. A description of the participants is shown in Table 2. The comparison with students who did not complete the final survey suggests that there are some patterns of selective nonresponse, in that respondents are slightly older, less likely to have a parent born in the Netherlands, perform better on the prescore, and more likely to have higher educated parents. The comparison between students in the personalized and non-personalized version supports that randomization was largely successful in generating two groups equal in expectation. A notable exception is the proportion of parents with education level “Middle,” which is higher for students working with the personalized version. However, this does not extrapolate into differences in the more extreme levels of education (i.e., “Low” and “High”), thereby yielding inconclusive support for any structural differences in parental education between the two experimental groups. Successful random assignment is further corroborated by the similarity in prescores, defined as the last test score for that subject prior to when the class started using the digital practising module. Importantly, whereas there has been some attrition with respect to participation in the postexperiment questionnaire, this does not seem to be related to the experimental condition students were assigned to. This makes it possible to draw causal inferences when comparing these two versions of the digital practising tool.

2.2 | Materials

Figure 2 visualizes the practising environment of students (screen on the right) and the dashboard used by teachers to monitor the practising behaviour of their students (screen on the left). Students log in with their personal access code and indicate which subject and chapter to practise with. They are then redirected to a screen similar to the one shown in Figure 2. The top bar shows the chosen subject and chapter, and below this bar information is displayed on the students practice performance (determined by practice duration and performance) and the number of exercises already completed. The practising environments of the personalized and non-personalized condition are identical.

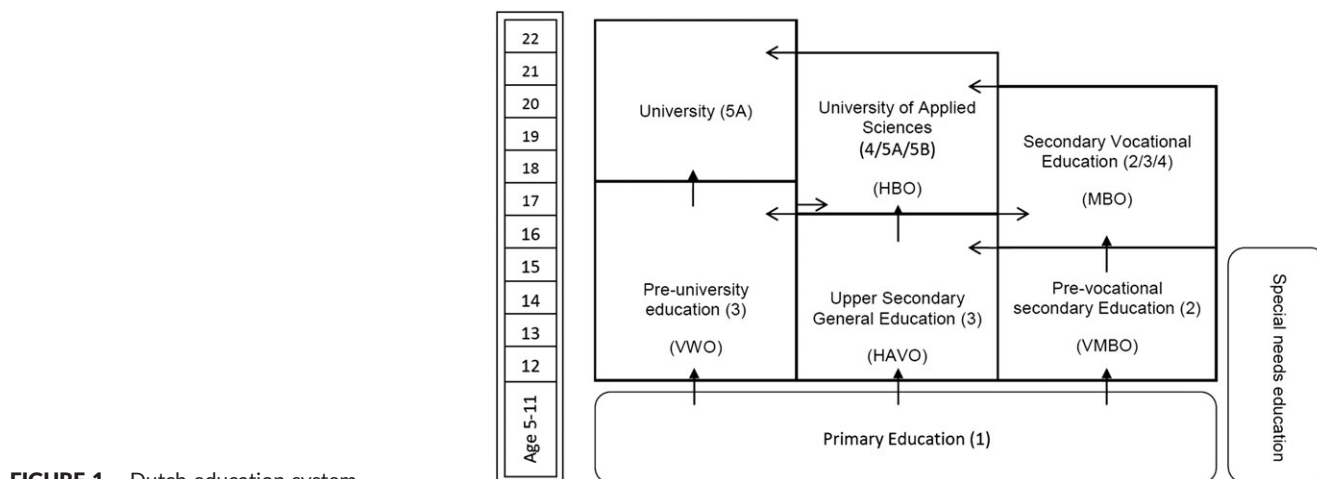


FIGURE 1 Dutch education system

TABLE 2 Student background characteristics by response on follow-up survey and personalization

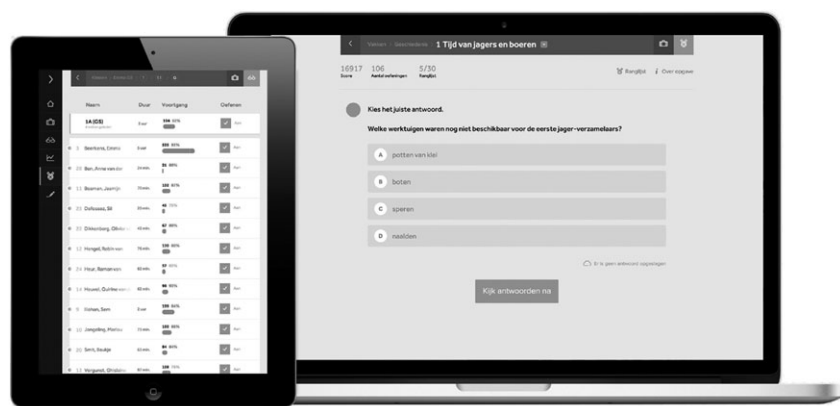
Boy	Follow-up survey comparison					Experimental group comparison				
	Respondents		Non-respondents		Diff. <i>p</i> value	Non-personalized		Personalized		Diff. <i>p</i> value
	Mean	Std. error	Mean	Std. error		Mean	Std. error	Mean	Std. error	
Age	13.756	0.057	13.445	0.061	0.002***	13.644	0.063	13.705	0.066	0.504
Home language; Dutch	0.938	0.011	0.920	0.021	0.423	0.936	0.014	0.931	0.014	0.822
Precore	6.506	0.074	6.152	0.142	0.019**	6.537	0.107	6.477	0.102	0.684
Mother										
Country of birth: Dutch	0.804	0.019	0.865	0.027	0.084*	0.837	0.021	0.804	0.023	0.292
Education: low	0.073	0.012	0.104	0.024	0.203	0.099	0.017	0.062	0.014	0.090
Education: middle	0.204	0.019	0.227	0.033	0.544	0.167	0.021	0.255	0.025	0.007***
Education: high	0.354	0.022	0.252	0.034	0.017**	0.321	0.026	0.333	0.027	0.735
Education: unknown	0.363	0.023	0.393	0.038	0.497	0.401	0.028	0.340	0.027	0.118
Father										
Country of birth: Dutch	0.820	0.018	0.877	0.026	0.090*	0.849	0.020	0.820	0.022	0.331
Education: low	0.081	0.013	0.104	0.024	0.374	0.106	0.017	0.069	0.014	0.102
Education: middle	0.163	0.017	0.209	0.032	0.186	0.147	0.020	0.203	0.023	0.071*
Education: high	0.389	0.023	0.294	0.036	0.031**	0.353	0.027	0.376	0.028	0.549
Education: unknown	0.360	0.023	0.374	0.038	0.754	0.388	0.028	0.340	0.027	0.216
N	455		163			312		306		

Note:

*Significant at 10% level (two sided).

**Significant at 5% level (two sided).

***Significant at 1% level (two sided).

**FIGURE 2** Digital practice environment

The dashboard for teachers displays information on the practice activity of the class and provides extensive information on the practising process of the individual students (i.e., practice time, number of exercises answered, and proportion exercises answered correctly [by knowledge type]). Teachers are free to choose to use this information when providing feedback to their students about their learning process.

Throughout the year, each practice session for a student is linked to the corresponding test for which the student is preparing. In general, students have a few weeks of classes, with corresponding homework, before a test is administered. After this, the next period starts, in which a new practice session is generated for each student in each of the participating subjects and with exercises linked to the next test for that particular subject. For each session, the student can voluntarily decide to take up the opportunity to participate or not (i.e., participation decision). During the year of the experiment, four

subjects participate in the experimental sample: Dutch, Biology, History, and Economics. For each session, it is monitored whether or not the student decides to practise (i.e., participation decision), together with the number of exercises, the success rate (i.e., proportion correct), total practice time (i.e., in minutes), and performance on the corresponding summative test (i.e., scale of 1–10).

At the end of the school year, all students engaged in a postquestionnaire, asking them about their experiences with, and perceptions towards, using the computerized practising software. This postexperiment questionnaire asked students to evaluate statements concerning whether the tasks were generally “too easy” (TE: “When using the tool, exercises I receive are often too easy”), “too difficult” (TD: “When using the tool, exercises I receive are often too difficult”), “interesting” (TV: “When using the tool, exercises I receive are interesting.”), and “useful” (TV_u: “The tool is useful in helping me to better understand the curriculum”). Ratings were performed on a 5-point Likert

scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Answer 3 defines the neutral reference category. Answers 1 and 2 are taken together, as well as 4 and 5, as to define whether a student expresses a particular disposition (positive or negative) towards computerized practising. This classification is used to empirically compare the group of students expressing a disposition towards the task to those responding neutral to the statement (i.e., reference group). An overview of the items and coding in the survey is provided in the Table 3.

The aforementioned task perceptions will be used in the analyses to gauge the extent to which the practising process is likely to promote or inhibit student engagement. Differences in these perceptions will be related empirically to two outcomes indicative of student engagement. First, every time a new practice period starts (i.e., in between two summative tests), a student will make the decision whether or not to take up the opportunity to voluntarily practise using the software. This participation decision will be influenced by whether previous experiences with the tool were generally positive or negative and a higher likelihood of participating in computerized practising is thus considered to be a sign of greater student engagement. Second, conditional on deciding to participate, students will use the tool longer (shorter), depending on whether they are more (less) engaged in the process. As such, a higher level of practice intensity (measured in minutes) in a particular session is considered to also be a sign of greater student engagement.

2.3 | Practising algorithms

The two versions of computerized practising considered in this study are different in terms of their underlying algorithms. Below these technical differences are described in more detail.

2.3.1 | Non-personalized practising

The algorithm underlying the non-personalized practising condition is intuitively illustrated in Figure 3. Each exercise, e , received a label which refers to the topic (t), the knowledge type level (k), and the level of difficulty (d). It follows that $e_n^{t,k,d}$ refers to exercise n in exercise subset $\{t, d, k\}$. The non-personalized version distinguishes between three difficulty interval levels and three knowledge types, which explains why Figure 3 shows nine exercise subsets per topic.

For each student, one exercise is randomly drawn from each subset $\{t, d, k\}$, beginning with $e^{1,1,1}$ (the upper-left red coloured square)

and ending with $e^{T,3,3}$ (the lower-right red coloured square). The selected exercises are ordered by, subsequently, topic, knowledge type, and difficulty level and returned one by one to the student in this specific order. Exercises are randomly selected and are an integrated part of the learning material that is offered in the school, such that unique practising sessions are generated for each student that are valid and representative for the upcoming summative tests.

Some subsets contain more exercises than other subsets, which are represented by the height (h) of the square around $e^{t,k,d}$ in Figure 3. The probability that a particular exercise is returned from subset $e^{t,k,d}$ equals $\left(\frac{1}{h^{t,k,d}}\right)/I$. This illustrates the importance of having enough content, because the probability that the algorithm offers a particular exercise to a student becomes larger as the amount of content decreases.

2.3.2 | Personalized practising

The algorithm of the personalized practising tool is visualized in Figure 4. The understanding level of student i and class c on topic t is determined by the proportion of correct answers given. Let P denote this proportion for topic t , then the relative understanding level of topic t for student i is $\left(\frac{P_{it}}{P_{ct}}\right)$. Higher values of $\frac{P_{it}}{P_{ct}}$ refer to better understanding levels of student i . Figure 4 shows that the width of each box (l_t) represents the inverse relative understanding for student i of a particular topic (i.e., $\frac{P_{it}}{P_{ct}}^{-1}$) and the total total length l is an indicator for the inverse of a student's overall relative understanding. It is convenient to take inverse values because larger l_t values indicate that topic t is less well understood and, as such, $\frac{l_t}{l}$ represents the probability that the algorithm will offer an exercise on topic t . The personalized practising version updates the relative understanding levels after each exercise practised and the probability that student i receives an exercise on topic t equals

$$\Pr(t) = \left(\frac{l_t}{l}\right)^{-1} = \left(\frac{\left(\frac{P_{it}}{P_{ct}}\right)^{-1}}{\sum_{t=1}^T \left(\frac{P_{it}}{P_{ct}}\right)^{-1}}\right)^{-1}. \quad (1)$$

After the algorithm has selected a topic t , it determines two parameters (μ and p). Parameter μ represents a topic-specific

TABLE 3 Student postexperiment questionnaire items and group labelling

		Student perception labels				
		Strongly disagree 1	Somewhat disagree 2	Neutral 3	Somewhat agree 4	Strongly agree 5
Task easiness	TE	“When using the tool, exercises I receive are often too easy”				
		$TE(+)$		Ref.	$TE(-)$	
Task difficulty	TD	“When using the tool, exercises I receive are often too difficult”				
		$TD(+)$		Ref.	$TD(-)$	
Task value—interest	TV_i	“When using the tool, exercises I receive are interesting.”				
		$TV_i(-)$		Ref.	$TV_i(+)$	
Task value—usefulness	TV_u	“When using the tool, exercises I receive are useful.”				
		$TV_u(-)$		Ref.	$TV_u(+)$	

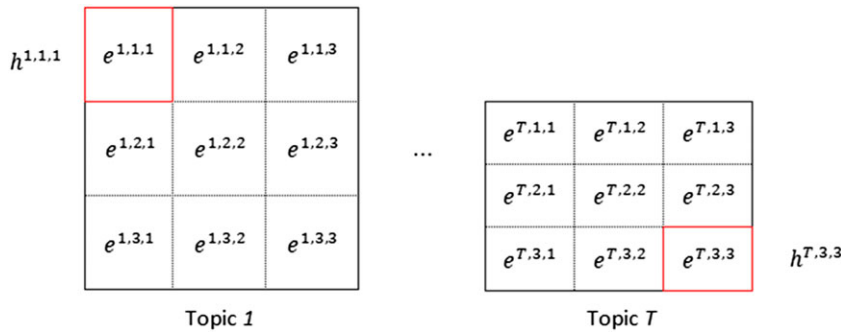


FIGURE 3 Intuition non-personalized algorithm [Colour figure can be viewed at wileyonlinelibrary.com]

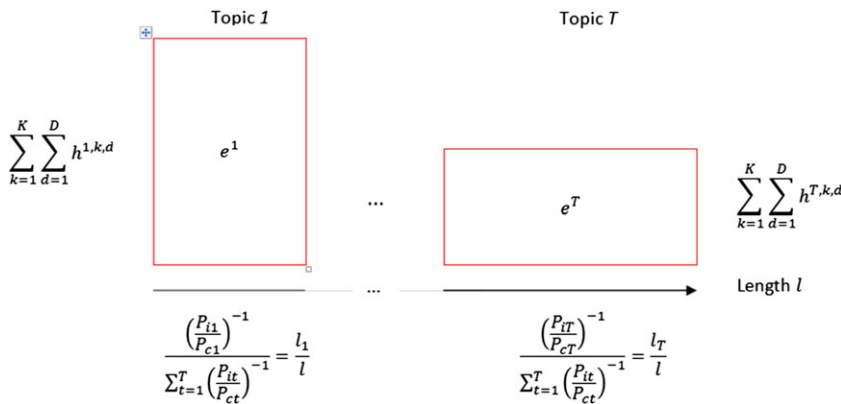


FIGURE 4 Intuition personalized algorithm [Colour figure can be viewed at wileyonlinelibrary.com]

threshold level which determines the knowledge type of the exercise that should be returned. This threshold level is set at 50%, which means that exercises of the current knowledge type are offered as long as $\left(\frac{P_{it}}{P_{ct}}\right)$ does not exceed 0.5. The algorithm switches and offers exercises of the next knowledge type if and only if $\left(\frac{P_{it}}{P_{ct}}\right) > 0.5$. Once the topic and the knowledge type are determined for a particular student, the algorithm returns an exercise for which holds that the understanding level of the average student $\left(\text{i.e., } \frac{1}{l} \sum_{i=1}^l P_{itk}\right)$ is closest to the

understanding level of the student (from above). In other words, an exercise will be returned that is slightly more difficult than the student's current understanding level.

The height of the square around e^t again refers to the available number of exercises on topic t , and the availability of sufficient content is of vital importance for the proper functioning of the personalization process. The figure illustrates a situation in which there are more exercises available for Topic 1 than for topic T $\left(\text{i.e., } \sum_{k=1}^K \sum_{d=1}^D h^{1,k,d} < \sum_{k=1}^K \sum_{d=1}^D h^{T,k,d}\right)$. At the same time, it indicates that

TABLE 4 Task easiness, difficulty and value, practice results and test scores—by practising condition

	Non-personalized		Personalized		Δ	Diff. p value
	Mean	SE	Mean	SE		
TE(-): task "easiness"—too easy	0.098	0.020	0.083	0.018	0.015	0.573
TE(0): task "easiness"—neutral	0.489	0.033	0.457	0.033	0.032	0.490
TE(+): task "easiness"—not too easy	0.400	0.033	0.457	0.033	-0.057	0.224
TD(-): task "difficulty"—too difficult	0.133	0.023	0.139	0.023	-0.006	0.857
TD(0): task "difficulty"—neutral	0.409	0.033	0.465	0.033	-0.056	0.227
TD(+): task "difficulty"—not too difficult	0.444	0.033	0.391	0.032	0.053	0.251
TV _i (-): task value "interesting"—negative	0.289	0.030	0.239	0.028	0.050	0.229
TV _i (0): task value "interesting"—neutral	0.373	0.032	0.426	0.033	-0.053	0.252
TV _i (+): task value "interesting"—positive	0.338	0.032	0.335	0.031	0.003	0.946
TV _u (-): task value "useful"—negative	0.302	0.031	0.309	0.031	-0.006	0.881
TV _u (0): task value "useful"—neutral	0.396	0.033	0.400	0.032	-0.004	0.923
TV _u (+): task value "useful"—positive	0.293	0.030	0.291	0.030	0.002	0.962
Total practising time (minutes)	116.6	9.438	132.7	11.89	-16.184	0.288
Proportion correct	0.766	0.005	0.689	0.005	0.077	0.000
Test score (rolling average)	6.262	0.094	6.307	0.087	-0.045	0.727
N	225		230			

the student's relative understanding of topic T is smaller than that of Topic 1 ($I_T < I_1$), which means that the algorithm will return an exercises of topic T with higher probability. The consequence of not having enough practising material would be that identical exercises are returned to the student frequently, such that the practising process is not so much personalized, but rather a process in which wrongly answered exercises are repeatedly returned. Furthermore, limited content would also make it difficult for the algorithm to closely match a student's individual level of relative understanding with an exercise to return. In this study, each subset contains many exercises for the algorithm to choose from (see also Footnote 1).

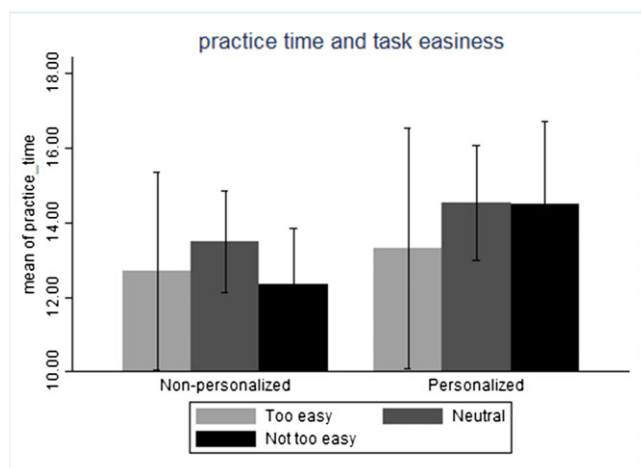


FIGURE 5 Task easiness and practice intensity [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS

As potential determinants, indications, and effects of student engagement, this study focuses on, respectively, task perceptions, practice behaviour, and student performance for two versions of computerized practising (i.e., non-personalized and personalized). Four hundred fifty-five students for which postquestionnaire answers are known are included in the results. These 455 students generated a total of 4,141 practice sessions ($M = 9.10$, $SD = 3.42$), resulting in a total of almost 57,000 min of computerized practising ($M = 124.7$, $SD = 162.3$), and with each session culminating in a test score for each student ($M = 6.47$, $SD = 1.78$) on a summative test corresponding to the content practised throughout the session and administered by the subject teacher. Throughout the school year, 454 students (99.8%) participated actively in at least one session or more, resulting in a total of 2,176 sessions (52.5%) for which the practice time was nonzero and positive. By the end of the school year, a total of 130,668 exercises ($M = 287.2$, $SD = 445.3$) were performed by these 454 students, after which they filled out the questionnaire asking them about their experiences and attitudes towards computerized practising.

Table 4 compares the student perceptions regarding the practising process, together with actual practising results and test scores (rolling averages) for both experimental conditions. Perceived task difficulty and task value of students in both practising conditions are not significantly different. Thus, at least on average, the personalized practising condition does not alter task perceptions, despite the statistically significant difference in actual success rate between the two experimental conditions. The 16-min gap in total practising time in favour of the personalized condition does not imply a statistically significant difference in practising intensity. The rolling average test

TABLE 5 Perceived task easiness (TE) related to participation decision, practice intensity, and test scores

	Participation		Practice intensity		Test score	
	1		2		3	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
TE(-)	-0.0487	(0.148)	-0.689	(3.481)	0.259	(0.193)
TE(+)	-0.0894	(0.073)	0.367	(2.987)	-0.169	(0.127)
Personalized	0.0582	(0.071)	3.740	(2.625)	0.0199	(0.116)
TE(-) × Personalized	0.0161	(0.197)	-2.550	(5.619)	-0.291	(0.360)
TE(+) × Personalized	0.197*	(0.101)	-2.871	(4.322)	-0.0345	(0.177)
Constant	-0.653**	(0.316)	1.385	(10.46)	4.751***	(0.620)
Prescore	Yes		Yes		Yes	
Background controls	Yes		Yes		Yes	
N	4,141		2,176		4,141	
# Student clusters	455		454		455	
pseudo- R^2	0.187		0.087		0.051	

Note. Dependent variable in probit Model 1 is whether or not a student engaged in practising in a particular session. Dependent variable in OLS-model 2 is how long (in minutes) participating students engaged in practising during a particular session. Dependent variable in Model 3 is the achieved test score on the corresponding test on scale [1–10]. Reference category is perceived task easiness—neutral. SEs are clustered at the student level. Prescore is the last test score for a particular subject prior to when the class actively participated in the experiment. Background control variables are boy, age, edu. level mother, edu. level father, Dutch language at home, and test number.

*Significant at 10% level (two sided).

**Significant at 5% level (two sided).

***Significant at 1% level (two sided).

scores between students who practised personalized and non-personalized are not significantly different either. Yet there is a substantial amount of variation in reported categories of task perceptions. This variation will be exploited empirically to examine if practice intensity is related to observed differences in task perceptions. For each dimension, the causal effect of being assigned to one of the two versions is evaluated.

In order to estimate the statistical relationships between these dispositions towards task perceptions and student outcomes, the following type of empirical model is estimated:

$$Y_{is} = \alpha + \beta_0 D_k(-)_i + \beta_1 D_k(+)_i + \beta_2 P_i + \beta_3 D_k(-)_i \cdot P_i + \beta_4 D_k(+)_i \cdot P_i + X_i' \gamma + \varepsilon_{is}, \quad (2)$$

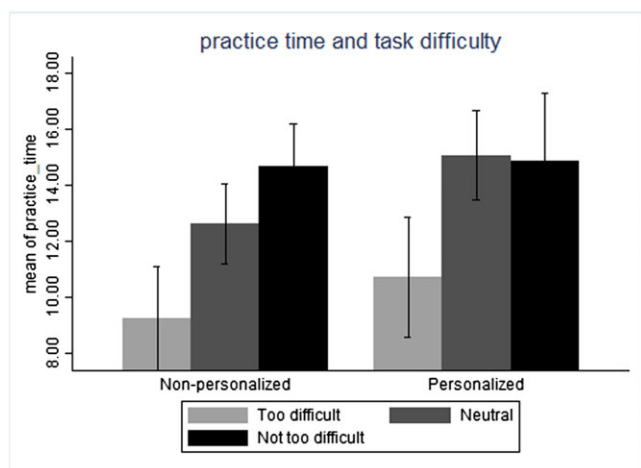


FIGURE 6 Task difficulty and practice intensity [Colour figure can be viewed at wileyonlinelibrary.com]

where Y_{is} represents the outcome (e.g., practice intensity) for student i in session s , $D_k(-)$ represents a dummy variable ($0 = no$, $1 = yes$) from the set $TE(-)$, $TD(-)$, $TV_i(-)$, $TV_u(-)$, indicating whether a student expresses a particular negative disposition towards practising (e.g., $TV_i(-)$ for task value “interesting”—negative), $D_k(+)$ represents a dummy variable ($0 = no$, $1 = yes$) from the set $TE(+)$, $TD(+)$, $TV_i(+)$, $TV_u(+)$, indicating whether a student expresses a particular positive disposition towards practising (e.g., $TV_i(+)$ for task value “interesting”—positive), and P_i indicates whether a student is assigned to the personalized practice condition ($0 = no$, $1 = yes$). Of particular interest with respect to the objectives for this paper are β_2 , which represents the causal effect of being assigned to the personalized condition, and β_3 (β_4) which indicates if there is a statistically significant difference in empirical association between the negative (positive) disposition and the outcome for students in the personalized condition. Finally, X_i' represents a set of student characteristics and test dummy variables, which are added in extended models for enhanced precision.

For student outcome “participation decision” ($0 = no$, $1 = yes$), the above equation is estimated using probit regression models, whereas for the student outcomes “practice intensity” (in minutes) and “test score” (range of 1–10) ordinary least squares regression models are estimated.

3.1 | Perceived task easiness and practice intensity

Students were asked to indicate if they generally experienced computerized practising as *too easy* [$TE(-)$], *neutral* [$TE(0)$], or as *not too easy* [$TE(+)$]. The relationship between perceived task easiness and practice intensity (measured as average practice time per session) is shown in

TABLE 6 Perceived task difficulty (TD) related to participation decision, practice intensity and test scores

	Participation		Practice intensity		Test score	
	1		2		3	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
$TD(-)$	-0.0805	(0.093)	-6.050*	(3.104)	-0.263	(0.290)
$TD(+)$	-0.0143	(0.081)	3.066	(2.955)	0.0207	(0.172)
Personalized	0.140*	(0.080)	4.237	(2.593)	-0.118	(0.170)
$TD(-) \times \text{Personalized}$	0.00368	(0.132)	-0.259	(4.551)	0.220	(0.355)
$TD(+) \times \text{Personalized}$	-0.000368	(0.111)	-4.749	(4.649)	0.137	(0.263)
Constant	-0.697**	(0.334)	4.441	(10.80)	5.412***	(1.092)
Prescore	Yes		Yes		Yes	
Background controls	Yes		Yes		Yes	
N	4,141		2,176		4,141	
# Student clusters	455		454		455	
pseudo- R^2	0.186		0.092		0.050	

Note. Dependent variable in probit Model 1 is whether or not a student engaged in practising in a particular session. Dependent variable in OLS-model 2 is how long (in minutes) participating students engaged in practising during a particular session. Dependent variable in Model 3 is the achieved test score on the corresponding test on scale [1–10]. Reference category is perceived task easiness—neutral. SEs are clustered at the student level. Prescore is the last test score for a particular subject prior to when the class actively participated in the experiment. Background control variables are boy, age, edu. level mother, edu. level father, Dutch language at home, and test number.

*Significant at 10% level (two sided).

**Significant at 5% level (two sided).

***Significant at 1% level (two sided).

Figure 5. In both conditions, there appears to be no clear difference between students who perceive computerized practising as either *too easy*, *neutral*, or *not too easy*.

The results reported in Table 5 Model 1 suggest that there is no clear relationship between perceived task easiness and the decision to participate in a particular practice session. For students who did practise, results for whether perceived task easiness affects practice intensity (Model 2) and student performance (Model 3) also indicate that perceived task easiness is not related to these outcomes either.

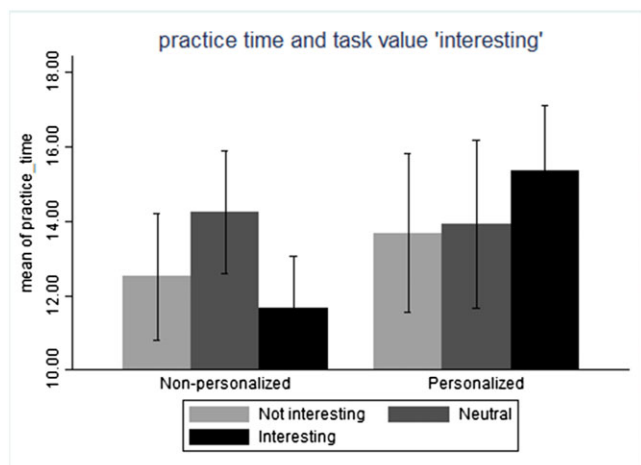


FIGURE 7 Task value “interesting” and practice intensity [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Perceived task difficulty and practice intensity

Students were also asked to indicate if they generally experienced computerized practising as too difficult [TD(-)], neutral, or as not too difficult [TD(+)]. The relationship between perceived task difficulty and practice intensity (measured as average practice time per session) is shown in Figure 6. In contrast to “task easiness,” the notion of “task difficulty” is systematically related to practice intensity. For students being assigned to the non-personalized practising conditions, it holds that practice intensity is positively related to not perceiving the task to be too difficult. The practice intensity for students assigned to the personalized practising is also lower if they perceive computerized practising as too difficult. However, in this condition, there appears to be no difference between students who perceive computerized practising as *neutral* and *not too difficult*. A potential explanation for this is that students assigned to personalized practising receive on average more difficult exercises (see Table 4). As a result, fewer students stated that practising sessions were perceived as *not too difficult*, as illustrated by the relatively large confidence intervals for this group. Those students experiencing practising as *not too difficult* receive on average more difficult exercises and this could cause convergence to the level of practice intensity of students who were neutral with respect to the difficulty level of computerized personalized practising.

The estimation results in Table 6 Model 1 suggest that there is no relationship between perceived task difficulty and participation in practising. For students who did participate, the estimation results in Model 2 suggest that practice intensity is reduced when the perceived task difficulty is too high and increased when the task is

TABLE 7 Perceived task value “interesting” (TV_i) related to participation decision, practice intensity, and test scores

	Participation		Practice intensity		Test score	
	1		2		3	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
$TV_i(-)$	0.109	(0.0791)	-3.528	(3.106)	-0.0751	(0.162)
$TV_i(+)$	-0.129	(0.0839)	-1.790	(3.142)	-0.0145	(0.133)
Personalized	0.0290	(0.0822)	0.517	(3.785)	0.0656	(0.121)
$TV_i(-) \times$ Personalized	-0.0131	(0.120)	4.842	(5.395)	0.0454	(0.220)
$TV_i(+) \times$ Personalized	0.312***	(0.113)	1.944	(4.849)	-0.283	(0.193)
Constant	-0.543*	(0.325)	4.229	(12.05)	4.732***	(0.618)
Prescore	Yes		Yes		Yes	
Background controls	Yes		Yes		Yes	
N	4,141		2,176		4,141	
# Student clusters	455		454		455	
pseudo- R^2	0.189		0.087		0.051	

Note. Dependent variable in probit Model 1 is whether or not a student engaged in practising in a particular session. Dependent variable in OLS-model 2 is how long (in minutes) participating students engaged in practising during a particular session. Dependent variable in Model 3 is the achieved test score on the corresponding test on scale [1–10]. Reference category is perceived task easiness—neutral. SEs are clustered at the student level. Prescore is the last test score for a particular subject prior to when the class actively participated in the experiment. Background control variables are boy, age, edu. level mother, edu. level father, Dutch language at home, and test number.

*Significant at 10% level (two sided).

**Significant at 5% level (two sided).

***Significant at 1% level (two sided).

considered not too difficult. However, only when task difficulty is considered too high, do the results appear to be statistically significantly different from the neutral category. Furthermore, the perception that the task of computerized practising is too difficult (or not), does not contribute to achieving higher (or lower) test scores (Model 3).

In comparing the results for task easiness and task difficulty, it can be concluded that only the perception that the task is too difficult seems to be relevant in explaining differences in student engagement with computerized practising, as indicated by observed lower levels of practice intensity.

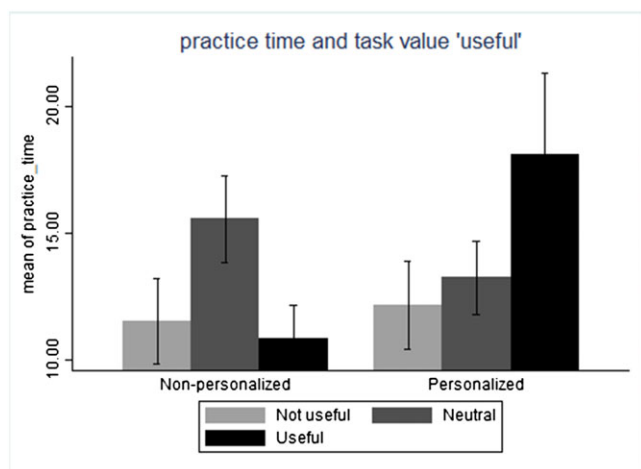


FIGURE 8 Task value “useful” and practice intensity [Colour figure can be viewed at wileyonlinelibrary.com]

3.3 | Task value and practice intensity

To infer elements of task value, students were asked to report whether they perceive computerized practising as *interesting* [$TV_i(+)$] or *useful* [$TV_u(+)$]. Figure 7 shows the relationship between practice intensity and the extent to which students perceive computerized practising as interesting. In general, the figure illustrates that practising intensity increases when students perceive practising as more interesting. Somewhat remarkable is that for students assigned to the non-personalized practising condition the practising intensity of students who consider this task to be interesting is not higher than that of students who find practising less interesting. The confidence intervals reveal that these differences are not statistically significant and, furthermore, these differences may be driven by differences in participation decisions and background characteristics.

Table 7 Model 1 shows that students who perceive practising as interesting are indeed more likely to participate in the practising process, but only if these students were assigned to the personalized practising process. Students in the personalized practising conditions receive exercises that better match their personal needs throughout the entire practising process, and this may induce them to practise. Students in the non-personalized practising condition may experience that the offered exercises do not match their personal practising needs and, as such, may be more likely to decide not to practise in upcoming sessions, even though they do find computerized practising interesting. Table 7 furthermore shows that perceiving computerized practising as interesting is not related to practising intensity (Model 2), nor test scores (Model 3), once students have decided to participate in the practising process.

TABLE 8 Perceived task value “useful” (TV_u) related to participation decision, practice intensity, and test scores

	Participation		Practice intensity		Test score	
	1		2		3	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
$TV_u(-)$	0.0519	(0.0817)	-2.726	(3.183)	-0.0854	(0.146)
$TV_u(+)$	0.0634	(0.0889)	-5.735*	(2.920)	-0.0146	(0.141)
Personalized	0.162*	(0.0849)	-2.166	(3.219)	-0.188	(0.139)
$TV_u(-) \times$ Personalized	-0.0706	(0.118)	3.001	(4.457)	0.314	(0.201)
$TV_u(+) \times$ Personalized	0.000782	(0.121)	11.12**	(5.060)	0.232	(0.211)
Constant	-0.749**	(0.327)	9.319	(10.96)	4.931***	(0.693)
Prescore	Yes		Yes		Yes	
Background controls	Yes		Yes		Yes	
N	4,141		2,176		4,141	
# Student clusters	455		454		455	
pseudo- R^2	0.186		0.092		0.049	

Note. Dependent variable in probit Model 1 is whether or not a student engaged in practising in a particular session. Dependent variable in OLS-model 2 is how long (in minutes) participating students engaged in practising during a particular session. Dependent variable in Model 3 is the achieved test score on the corresponding test on scale [1–10]. Reference category is perceived task easiness—neutral. SEs are clustered at the student level. Prescore is the last test score for a particular subject prior to when the class actively participated in the experiment. Background control variables are boy, age, edu. level mother, edu. level father, Dutch language at home, and test number.

*Significant at 10% level (two sided).

**Significant at 10% level (two sided).

***Significant at 10% level (two sided).

Figure 8 visualizes the relationship between practice intensity and whether students perceive computerized practising to be useful. The results are very similar to those in Figure 7, in the sense that practising intensity increases when students perceive practising as useful, but only in the personalized practising condition. The estimation results in Table 8 Model 1 show that the participation decision of students to practise is not driven by their perceived judgment on whether practising is useful. Practice intensity of students who perceive the practising process as useful is significantly lower for students assigned to the non-personalized practising condition, but higher for students assigned to the personalized practising condition (Model 2). Again, these results do not translate into differential patterns in test scores (Model 3).

In short, the results for task value interest and usefulness point out that both have the potential to promote student engagement, albeit in different ways and only when students work in a personalized

practising environment. Students who perceive the task as interesting are more likely to take up the opportunity to practise when assigned to the personalized condition, but conditional on this effect, no differences in practice intensity are observed. For students who perceive the task as useful, the potential to improve student engagement in the personalized condition is not so much driven by a higher likelihood to participate, but by the fact that, once students decide to practise, they will do so for a longer period of time.

3.4 | Practice intensity and task perceptions evolving over time

In the previous analyses, average practising intensities were related to the task perceptions measured at the end of the school year. However, the observed patterns of practice intensity over the school year

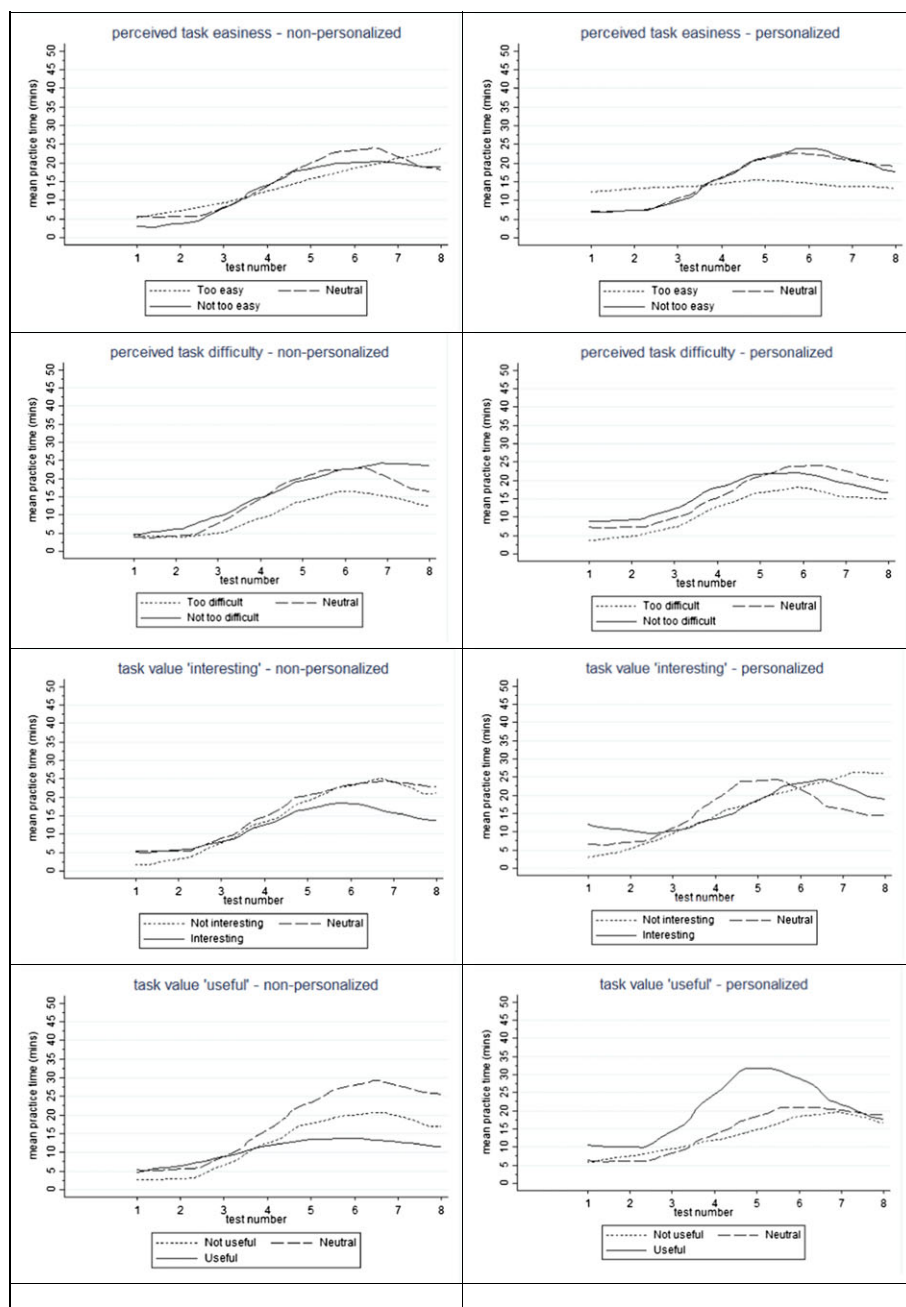


FIGURE 9 Task perceptions related to practice intensity over time [Colour figure can be viewed at wileyonlinelibrary.com]

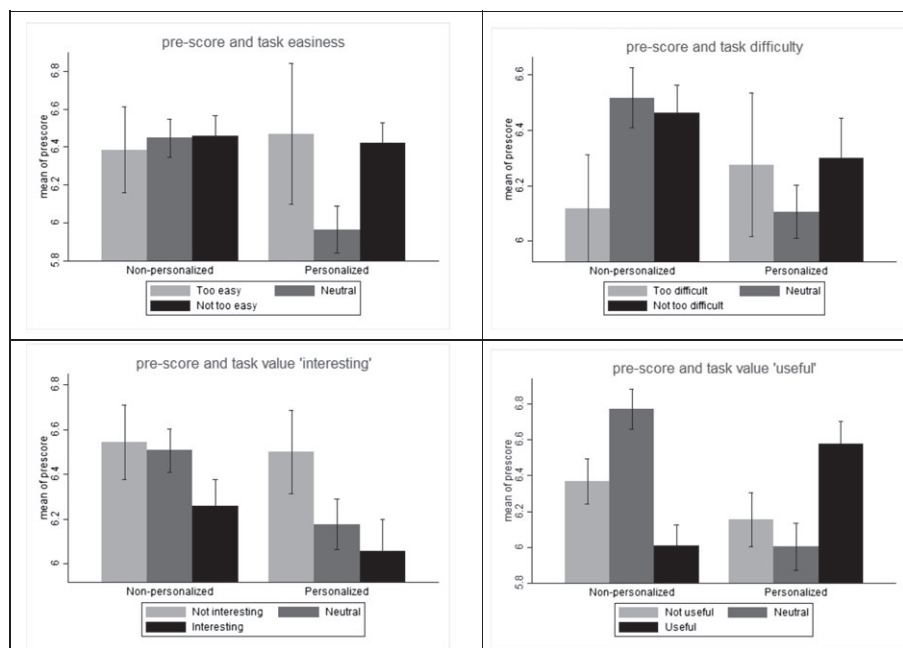


FIGURE 10 Heterogeneous effects of personalization by ability

is dynamic and, as a result, differential patterns in practising time may emerge over the course of the school year and across different levels of perceptions. This gives insights into the potential for student engagement to either improve or deteriorate as experience with the task increases. Figures 9 shows these patterns for the various task perceptions considered throughout this study.

As to be expected, early on in the year (i.e., in the absence of experience with the task), practice intensity is still largely similar for students who end up differing considerably in terms of task perceptions. Practice intensity initially increases with test number, which indicates that students tend to practise more over time throughout the school year after which it levels off. There are some notable differences in patterns between students belonging to different categories of task perceptions. In general, the observed increase in practice intensity is smallest for students who perceived practising as too difficult or too easy, which is in line with theoretical predictions. Furthermore, the differential results with respect to task value and assigned condition presented in the previous section are confirmed in these dynamic patterns and seem to emerge in the second half of the year (i.e., after test Number 4).

3.5 | Effects of personalization on task perceptions by ability

Given that implications of technological interventions can differ importantly with respect to ability, the previous results are separated by prescore. This prescore is obtained from the last test prior to when the class started using the digital practising module for that particular subject and can be used as a proxy for subject-specific ability. Several differences emerge between the two assignment conditions, as presented in Figure 10. The non-personalized version is considered to be too difficult particularly by students with a relative low prescore. The personalized version is designed to adapt and expose better students to more difficult exercises and weaker students to less difficult

exercises. This is confirmed in that there is no statistically significant difference in test scores between students who consider the exercises to be (not) too difficult in the personalized condition. With respect to task value, a general pattern is observed for both conditions in that students who find the tool interesting are generally students of lower ability (cf. Cheung & Slavin, 2013). However, a notable difference emerges with respect to whether exercises are considered to be useful. In the non-personalized version, students evaluating the tool as not useful have somewhat lower prescores, which is even more the case for students considering the tool to be useful. Instead, for students assigned to the personalized version, it is particularly the relatively high-performing students who evaluate the tool as useful. In short, the non-personalized version particularly appealed as useful to relatively low-ability students, whereas the reverse is found for the personalized condition.

4 | CONCLUSION AND DISCUSSION

With the objective to improve student engagement, this study examines the relationship between the effort students exert in computerized practising and several task perceptions (i.e., easiness, difficulty, interest, and usefulness). In addition, it provides causal evidence for whether, and which, students exert more effort when the offered practising process is personalized. In this personalized version, topics which the student finds relatively difficult are offered with higher probability, knowledge types are contingent on whether mastery has been displayed and the difficulty level of exercises is matched to a student's current level of understanding. The empirical analyses draw on information collected for 455 secondary school students throughout a school year. Students are randomly assigned to a personalized version, designed to address their individual learning needs, or to a non-personalized version, designed to return sequences of exercises

that are representative for the corresponding summative assessment. During this period, the academic performance and practising intensity of students was monitored for four different subjects. At the end of the school year, students' task perceptions regarding the computerized practising process were collected by means of a questionnaire.

The results show that there is significant variation in perceived task difficulty and the two dimensions of task value considered (i.e., interest and usefulness). Students randomly assigned to work with this personalized version have on average a lower success rate when practising. However, this does not translate to differential rates of average practice intensity, nor to differences in perceived attitudes towards practising. Furthermore, academic performance on summative tests is similar for both groups.

Regarding the potential detrimental effects on motivation to practise of offering students tasks that are either too difficult or under challenging (Ciampa, 2014; Tomlinson et al., 2003). The findings reveal that—in the context of this study—only the former is related to reduce practice intensity. That is, when students perceive the practising process as too easy, this does not translate to differences in student effort. Yet whether students report that the task was generally too difficult, this is empirically associated with lower levels of practice intensity (measured in minutes spent per session). This should be acknowledged when designing and implementing modern technologies in schools for their ability to effectively differentiate content and allow for sustained self-paced learning (Ciampa, 2014). The results reported here support the notion that self-esteem maintenance motives are critical determinants of task value for adolescent learners and that their level of sustained motivation declines beyond a certain level of perceived task difficulty (Eccles & Wigfield, 1995). Keeping students actively engaged in using a practising tool should thus offer content that is generally within reach of the individual student.

The results for perceived task interest and usefulness point out that both have the potential to promote student engagement, as predicted by Expectancy-Value Theory (Eccles & Wigfield, 1995), albeit in specific ways and only when students work in a personalized practising environment. Whether the task is perceived as interesting is related to a higher propensity to practise when a student is assigned to the personalized condition, but—conditional on this decision to practise—no differences in practice intensity are observed. Yet when the task is considered to be useful, results are different in that the potential to improve student engagement in the personalized condition is through higher levels of practice intensity instead. These implications can be summarized by stating that, in the context of this study, task interest can help explain a student's decision to engage in the process initially, whereas perceived usefulness is a more important determinant in explaining prolonged levels of effort sustainment during the actual practising process. Therefore, in the process of developing a digital tool for students, such theoretical underpinnings should be properly acknowledged in order for it to be effective for specifically those educational outcomes it aspires to improve (Jeng et al., 2010).

Exploring how practice intensity evolved over time furthermore reveals that students first need to gain experience with the task, before differences in perceived attitudes emerge and correlate to corresponding differences in practice intensity. When separated by ability, results regarding task perceptions are markedly different for the

two versions of computerized practising. In the non-personalized condition, students with lower prescores value practising as useful and as too difficult; whereas in the personalized condition, students with higher prescores value processing as useful. One plausible interpretation for this is that students of lower ability consider practising as useful when they do not get disengaged due to a difficulty level that is too high (Eccles & Wigfield, 1995) and if the process closely resembles the summative test for which they are preparing, whereas higher ability students attribute more usefulness value when practising is relatively challenging and more directly addressing their personal learning needs. The implication of this for the development of adaptive software thus seems to be that students can vary importantly in the relative weights they assign to different dimensions of usefulness. When implemented as a relatively low-stakes activity, as is often the case in educational practice, the results confirm the importance of perceived usefulness in predicting student effort (see also Cole et al., 2008). They also seem to indicate that personalizing the practice process is more likely to engage students if differences in individual characteristics (e.g., low vs. high ability), preferences (e.g., test preparation vs. addressing learning needs), and learning styles (Truong, 2016) are properly acknowledged. This reinforces the claim that the power of using technology is premised on the notion that such tools take into account the existing pedagogy (Ciampa, 2014) and educational setting in which they are implemented.

One limitation of this study is that student perceptions regarding practising were retrieved from nonvalidated questionnaires and only at the end of the experiment. For a better assessment of the relationship between expectancy task value and engagement with computerized practising, these perceptions should ideally be measured more frequently (i.e., throughout the year and during practising), more reliably (i.e., multiple items per value construct), and more comprehensively (i.e., also including expectancy and ability perceptions). In addition, it would be important to also collect measures of the perceived cost of practising. Cost in this context has been shown to be a multidimensional construct (Flake, Barron, Hulleman, McCoach, & Welsh, 2015) dealing with what is invested, required or given up in order to engage in the task, and seems particularly relevant for explaining student motivation in low-stakes contexts (Wise & DeMars, 2005). Recent evidence based on middle-school students confirms that cost can indeed be considered, and measured as, a distinct type of domain-specific motivation (Kosovich, Hulleman, Barron, & Getty, 2015). Another limitation is that summative tests were designed and administered by each subject teacher, making it impossible to link items from the practising process to items on the test and to infer whether the practising content strongly resembled that of the summative test. The latter might help explain why this study found no significant results with respect to personalized practising and student performance.

For future research, it is furthermore important to acknowledge that one explanation for the suboptimal results of current computerized personalized practising tools (see: OECD, 2015) may well be that adaptive processes are generally offered in a one-size-fits-all approach, as was the case in this study. The results presented in this paper with respect to ability suggest that students would assign more task value to computerized practising if the process would properly take into account the heterogeneity of the student population. It can

be argued that there are infinite ways to personalize the learning process and optimal adaptation of the content offered to a heterogeneous group of learners cannot be realized using only a single algorithm. Optimally, a system continuously evaluates for each student the algorithm which best accommodates the individual learning needs, goals, and preferences. As a result, it may well be that a single tool deploys multiple algorithms in order to continuously adapt the practising process to the individual learner.

ORCID

Ilja Cornelisz  <http://orcid.org/0000-0002-9948-7861>

REFERENCES

- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., ... Wittrock, M. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Ashcraft, M. H., Krause, J. A., Hopko, D. R., Berch, D., & Mazzocco, M. (2007). *Is math anxiety a mathematical learning disability. Why is math so hard for some children.* (pp. 329–348).
- Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review*, 64, 359–372.
- Bloom, B. S., & College of Confidential., and University Examiners (1956). In D. McKay (Ed.), *Taxonomy of educational objectives, volume 1*. New York.
- Bulman, G. and Fairlie, R. W. (2016). Technology and education: Computers, software, and the internet. Technical report, National Bureau of Economic Research.
- Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3), 198–215.
- Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.
- Ciampa, K. (2014). Learning in a mobile age: An investigation of student motivation. *Journal of Computer Assisted Learning*, 30(1), 82–96.
- Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology*, 33(4), 609–624.
- Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.
- Demski, J. (2012). This time it's personal: True student-centered learning has a lot of support from education leaders, but it can't really happen without all the right technology infrastructure to drive it. and the technology just may be ready to deliver on its promise. *Technological Horizons In Education Journal*, 39(1), 32.
- Eccles, J. S., Adler, T. F., Futterman, R., Go, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). *Achievement and achievement motivation, chapter expectancies, values, and academic behaviors.* (pp. 75–146). San Francisco, CA: W. H. Freeman.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21(3), 215–225.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132.
- Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, 41, 232–244.
- Hassler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning*, 32(2), 139–156.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Jeng, Y.-L., Wu, T.-T., Huang, Y.-M., Tan, Q., & Yang, S. J. H. (2010). The add-on impact of mobile applications in learning strategies: A review study. *Educational Technology & Society*, 13(3), 3–11.
- Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence*, 35(5–6), 790–816.
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218.
- Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265–299.
- Kulik, J. A. (2003). Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say. Citeseer.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449–521.
- National Research Council (2003). *Engaging schools: Fostering high school students' motivation to learn*. National Academies Press.
- Newmann, F. M. (1992). *Student engagement and achievement in American secondary schools*. ERIC.
- OECD (2015). *Students, computers and learning: Making the connection*. PISA.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Sherhoff, D. J., Csikszentmihalyi, M., Schneider, B., & Sherhoff, E. S. (2014). Student engagement in high school classrooms from the perspective of flow theory. In *Applications of flow in human development and education* (pp. 475–494). Springer.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57(2), 175–213.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15–21.
- Slavin, R. E. (2004). Education research can and must address "what works" questions. *Educational Researcher*, 33(1), 27–28.
- Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., ... Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, 27(2–3), 119–145.
- Truong, H. M. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, 55, 1185–1193.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, 23(3), 34–41.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
- Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10(1), 1–17.

How to cite this article: Cornelisz I, van Klaveren C. Student engagement with computerized practising: Ability, task value, and difficulty perceptions. *J Comput Assist Learn*. 2018;34: 828–842. <https://doi.org/10.1111/jcal.12292>