Improved Performances and Motivation in Intelligent Tutoring Systems: Combining Machine Learning and Learner Choice

Benjamin Clément, Hélène Sauzéon, Didier Roy, Pierre-Yves Oudeyer

Improved Performances and Motivation in Intelligent Tutoring Systems: Combining Machine Learning and Learner Choice.

Benjamin Clément[1,3,×], **Hélène Sauzéon**[1,2,+], **Didier Roy**[1], **and Pierre-Yves Oudeyer**[1,+,×]

[1]Inria, FLOWERS team, Talence, 33405, France
[2]Université de Bordeaux, BPH lab, Bordeaux, 33076, France
[3]EvidenceB, Paris, 75018, France
[×]corresponding authors
[+]these authors equally supervised this work

ABSTRACT

Large class sizes pose challenges to personalized learning in schools, which educational technologies, especially intelligent tutoring systems, aim to address. In this context, the ZPDES algorithm, based on the Learning Progress Hypothesis (LPH) and multi-armed bandit AI techniques, sequences exercises that maximize learning progress for each student. Previous field studies showed its learning efficacy compared to a hand-designed curriculum. However, its motivational impact was not assessed. Also, ZPDES did not allow students to express choices: this limitation in agency conflicts with LPH as a model of curiosity-driven learning. We study here how introducing choice (on dimensions orthogonal to exercise difficulty, acting as gamification) impacts both learning efficiency and motivation.
We present an extensive field study (265 7-8 years old children, RCT design) showing that ZPDES indeed improves learning performance but also produces a positive learning experience. Combining choice with ZPDES triggers intrinsic motivation and reinforces the learning effectiveness of the LP-based personalization. Conversely, adding choice possibilities to a hand-designed linear pedagogical paths produces deleterious effects on learning. Thus, the intrinsic motivation elicited by choice (gamification) is beneficial only if the curriculum is personalized efficiently for the learner. This deserves attention due to increased use of playful features in educational technologies.

1 Introduction

A key challenge of 21st century schools is to make students active and engaged in their education with the difficulty of dealing with a wide diversity of students' abilities and motivations for learning. The growing research on personalized or individualized education as well as on active teaching testifies to this huge societal need targeting the equality of opportunities at school for all[1].

The evidence-based assets of personalized learning over one-size-fits-all educational approaches are today well documented[2,3]. As classroom sizes are still high, it is difficult for teachers to set up individualized teaching paths, which is why high expectations are placed on Educational Technologies (ET) to automate them and support teachers in their missions.

The exploitation of artificial intelligence and digital systems has then become a crucial question to improve ET[4] and enable forms of adaptivity and personalization, leading to the development of Intelligent Tutoring Systems (ITS). This aims to make education more effective and accessible for the large diversity of students and as a way to provide useful objective metrics on learning[5–7].

Among the ITS field, there has been several approaches to optimize, personalize and adapt ITS to learners in order to enhance access to quality learning experience for all learners. Adaptation in learning technologies can be described as a structure with three main components[8].Firstly, adaptation to the instruction source, i.e. to what it will be adapted, such as the learner learning style[9,10], knowledge[11] or preferences[12]. Secondly, the target of the adaptive instruction, i.e. what will be adapted, such as the content[9] or the presentation[13]. Thirdly, the adaptive component generating a pathway between the two first components, i.e. how to adapt a Target to a Source, such as rule-based systems[9] or Bayesian-networks[14].

This last component is the engine generating a curriculum of training activities for learners in ITS. Thus, to be able to adapt the content to the learner, all ITS are most often implicitly or explicitly based on the concepts of the zone of proximal development (ZPD)[15] and the state of Flow[16]. These are well-known concepts in developmental psychology, and have inspired many models of learning for education (e.g., Cognitive load theory[17]). Following them, many ITS aim to offer the learner

pedagogical activities that are neither too difficult nor too easy with regard to their abilities, so that they can be engaged and progress in their acquisitions without being anxious or bored during the process. ITS can also propose activities the learner can not solve alone but will be able to solve with hints or with the teacher's help. From it, the ITS can be divided in two main design categories for managing learning curricula[1] : The former, namely the "linear design" involves that all learners follow a single path although at different speeds or with a different number of attempts. The latter, called "branched-paths design", enables each learner to follow a specific path according to their their own need.

Recently, with the growing interest in the phenomenon of curiosity[18], these concepts have been revisited in the Learning Progress (LP) model[19,20]. In this contemporary reward-learning model, curiosity-based intrinsic motivation and learning progress are linked by a virtuous loop: the child learns better on tasks for which she is interested and intrinsically motivated and, in return, the learning progress yielded generates an internal reward that stimulates his intrinsic motivation to continue acquiring knowledge for its own sake[21]; (see also Self-determination theory[22]). In other words, this model stresses the personal factors in learning, where individual LP contributes to both learner's motivation and self-organisation of its active exploration of learning tasks[23]. Furthermore, machine learning research has shown that using LP to automatically generate learning curricula, by sampling tasks with maximal expected LP, is a powerful heuristic that leads to sample efficient learning of skills and world models[24–30]. In other words, when one aims to maximize long-term learning outcomes on a variety of tasks, LP can be used as an efficient proximal heuristic for sampling learning activities. Thus, the LP model argues that LP-based learning curricula shall lead both to intrinsically motivating and efficient long-term learning.

In this context, the ZPDES algorithm (Zone of Proximal Development and Empirical Success) has been proposed to be used as a new LP-based activity manager in ITS[31].

Leveraging multi-armed bandit algorithms[32,33], it integrates an expert knowledge rule-based system and exploits the LP to select the activities which present the highest learning value for the student progress. Basically, ZPDES algorithm uses the student success rate to dynamically computes the best activity set (ZPD) and ponders the activities from LP to select them stochastically. The general idea is illustrated in figure 1 and shows the evolution of the possible activities available to the student as compared to what happens with a predefined linear sequence (Predef). ZPDES has been evaluated as an efficient algorithm for adaptive generation of various learning paths for real students[34], robust to heterogeneous populations as shown by complementary systematic experiments with simulated learners[35].

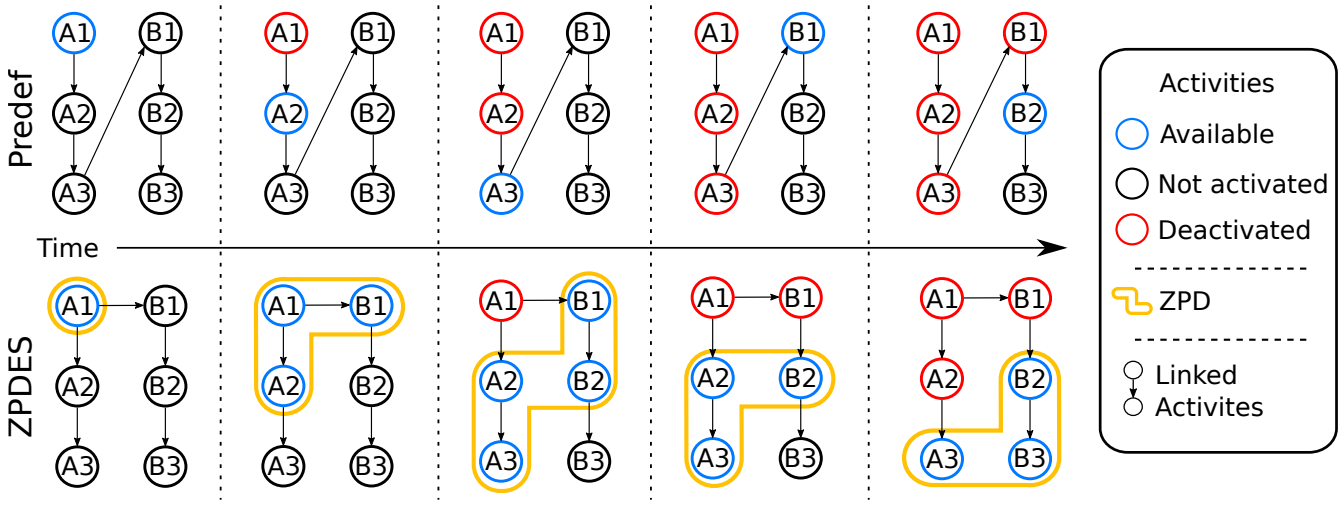


Figure 1. The space of available activities always contain only one activity in the predefined linear sequence (Predef) while the space expends over time with ZPDES to allow a diversity of exploration and find the best activities for the learner.

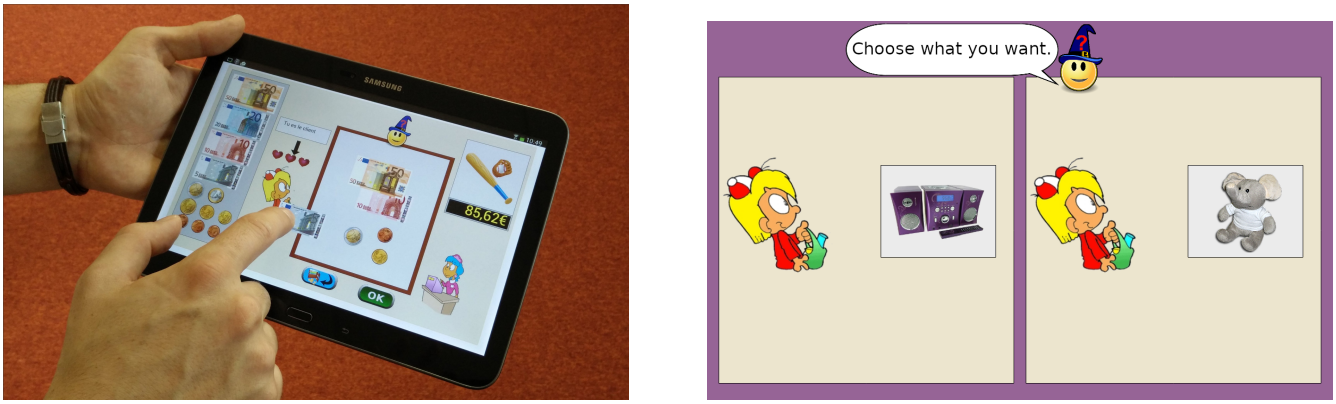
In summary, the automated personalization performed by ZPDES aims to lead learners into activities providing maximal learning progress, resulting in individualized learning paths. Yet, this approach had so far two major limitations. First, while its associated objective was to enhance intrinsic motivation, the motivational impact of ZPDES was not studies so far on human learners. Second, ZPDES does not lead the learner to actively make decisions related to the learning path although the theoretical LP model[19] encompasses self-decision making. Indeed, allowing self-decisions can boost the sense of agency and have a positive motivational impact, and be an efficient vector of performance[36–38]. One original argument for this design choice for ZPDES relied on findings in educational psychology, in particular those related to the self-regulation model of decision making[39] revealing that decision making can be biased and error prone, particularly in children[40] which were a priority target of this approach. A child can lack a given resource required to make an adaptive decision (e.g., lacking adequate knowledge) or

some factors could constrain the person's ability to carry out decision-making processes (e.g., under-/over-estimation of learner regarding his learning progress)[39,40].

A simple way to overcome potential decision making failures, while promoting the learner's intrinsic motivation to perform the activity, is to combine the use of ZPDES (to control the curriculum difficulty and variety) with offering choice possibilities limited to dimensions that are orthogonal to the learning complexity (e.g. here, choice of visual objects on which to do a math exercise as shown in Fig. 2b): this is the principle of ZCO[1] system introduced in this paper. Indeed, the subjective value of a task influences academic performance[41,42] and allowing young students (3rd through 9th-grade students[43]) to express their interest or preferences stimulates their intrinsic motivation and thus could be a booster for ITS effectiveness. An open question we address here is to understand the relative contributions and interactions of LP-based curriculum personalization and choice over both learning efficiency and motivation.

From the overall data, our first contribution is to show that ZPDES allows human learners to be more motivated and to learn better than a linear design based activity manager (which was made in collaboration with a pedagogical expert in maths teaching) named "Predefined sequence" (Predef), confirming the theoretical LP model and the relationships between learning progress and learner's intrinsic motivations. The second contribution is to show the synergistic effect between ZPDES and the ability given to children to express some choices (i.e. to make some decisions which is also a general form of gamification) on both learning performance and motivation. On the contrary, we show that giving children the ability to express some choices in the predefined sequence approach lowers the learning performance. Thus, the effect of choice on learning performance is contextual, and is here positive only for personalized learning curricula.

These two contribution are the result of a field study conducted according to an randomized control trial (RCT) design. Indeed, several systematic reviews reported promising or even positive results on the value-added of ITS[44–46], while pinpointing methodological limitations of this new empirical field (no control group, no initial group equivalence, no pre- and post-intervention measurements, etc.,[47]) and the great variability of the ITS designs making it difficult to identify which of the ITS features are critical for successful personalized learning[1,48,49].



(a) Kidlearn software on tablet

(b) Contextual Choice given to the student

Figure 2. Kidlearn software user interface

This RCT, approved by Inria COERLE ethical committee, involved 265 children, from 24 classes of 11 primary schools of the Bordeaux school district. The software on which children studied during this RCT, named Kidlearn ITS[34], was designed and developed specifically for this kind of experiment. This ITS aims to teach basic mathematics for children aged 7 years old through manipulation of money bills and coins (number decomposition, addition, subtraction of integers and decimals), and has been aligned to official pedagogical objectives on this topic in the national French education system. Figure 2a shows Kidlearn interface on a tablet the students use. They either play the role of the client or the merchant and need to compose the correct amount of money to either pay or give the change by dragging and dropping the bills and coins on the left side. You can refer to section 5.5 for more details about the pedagogical scenario and interface description. We compared 4 versions of the KidLearn ITS[34]; the Predefined sequence without (Predef) or with (PCO) learner decision, and ZPDES condition without (ZPDES) or with (ZCO) learner decision.

[1] ZCO stands for Zone of proximal development with ChOice

2 Results

The data presented here involves 265 children (Predef: 62, PCO: 59, ZPDES: 76, ZCO: 68) from schools of the Bordeaux school district. The experiment consisted of four sessions per class over two successive week (two sessions per week with at least one day break between each sessions). During the sessions, students interact with a tablet and either answer questionnaires (pre/post test, motivation questionnaires, etc) or work on Kidlearn ITS for 30 minutes (without planned interruption). The detailed experimental schedule and overall setup are presented in section 5.6.

To study the impact of LP-based personalizing and self-decision-making on student's learning performance and motivation, we use ZPDES as the LP-based algorithm (Sec. 5.2) and a predefined sequence following a "linear design" (called "Predef" described in Sec. 5.3) as a baseline. This predefined sequence is implemented as a series of activities in which the student must have 75% success over 4 activities of the same type to pass to the next activity type. The impact of self-decision making, an assumed in the LP-model, is also studied. It takes the form of a contextual choice given to the student consisting in choosing the visual objects presented during the exercise, thus the choice does not impact the difficulty of the exercise (Sec. 5.4). This leads to the comparison of four experimental conditions, two conditions without self-decision-making: ZPDES and Predef; two conditions with it: ZCO (ZPDES with Choice of Object) and PCO (Predef with Choice of Object).

We first check the differences in curricula, i.e the student progression in the Kidlearn app scenario, between each conditions. To grasp these differences, we compare the learning activities they reach and achieve during training as well as activity space evolution through time (Sec. 2.1). Differences in curricula raise the question of the learning effectiveness of each conditions (Sec. 2.2 and their impact on learning experience and motivation (Sec. 2.3). The learning effectiveness is evaluated through comparison between pre- and post-test results, while an emotional scale is used to assess the emotional valence of learning experience and the motivation is evaluated through Vallerand's questionnaire[50,51] (see Sec. 5.6.3 for more information on measures). Then, the relations between LP-based personalization and subsequent learning performance and motivation are analysed (Sec. 2.4). Finally, we check if individual characteristics modulate the impact of LP-based personalization of (Sec. 2.5).

2.1 How do LP-based personalized curricula differ from hand-designed ones?

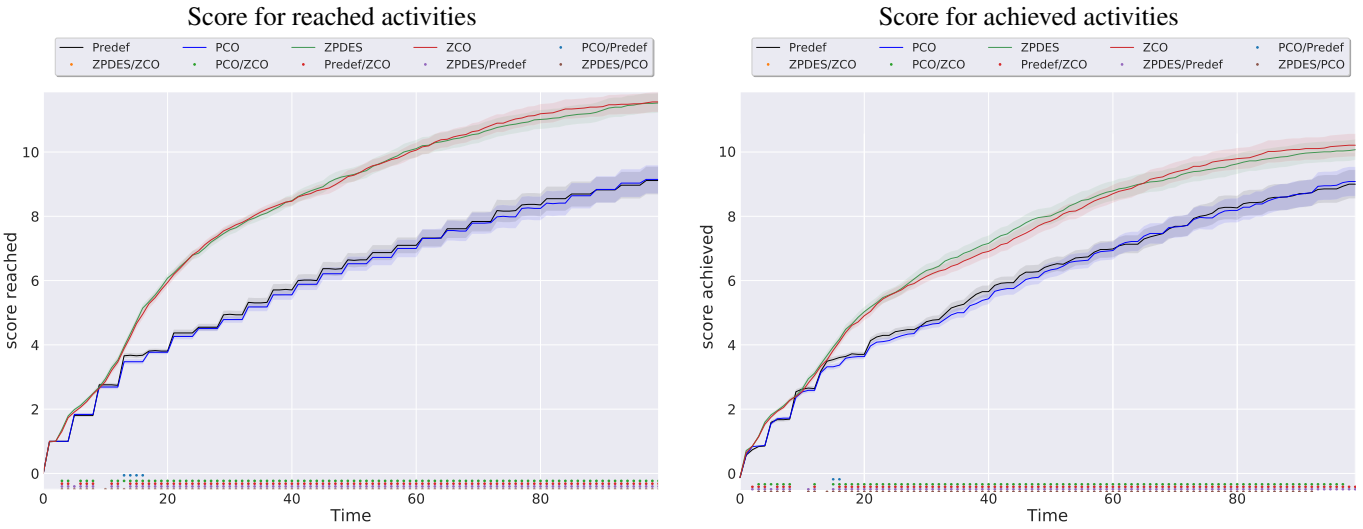


Figure 3. After 15 steps, students working with ZCO and ZPDES are reaching and succeeding more difficult and diverse activities than student working with PCO and Predef conditions. The curves represent the average score over all students for one condition. Activities "reached" are those that students have practiced, while activities "achieved" are those that students both practiced and succeeded. The shaded area represent the standard error of the mean. Colored points indicate if the score differences are significant two by two for each time step through t-test procedure.

It is not always trivial to grasp meaningful statistical differences between curricula over a population of student. Thereby, in order to be able to compare globally and quantitatively the curricula generated in each conditions, an "Activity score" has been designed (defined in Sec. 5.6.3). This score represents the level of difficulty either reached or achieved (i.e. reached *and* succeeded) by the students for each type of activity. In other words, we use this score as a proxy to understand how the student activity space generally evolves across time for each conditions.

Figure 3 shows the evolution of the average "Activity Score" for each condition across time. The shaded area represents the standard error of the mean and colored points indicate if the score differences are significant two by two through t-tests (as

presented in the top legend of the figure).

We can observe that before 15 steps, there is not much differences between the conditions. But, after 15 steps, the scores of the students working with ZPDES and ZCO grow faster than the scores of the students working with Predef and PCO. This means student working with ZPDES and ZCO are doing more diverse and difficult activities through time (and succeed in more diverse and difficult activities).

Also, giving a contextual choice does not seem to affect the activities done by the students due to the fact that ZPDES and ZCO scores are really similar, as it is for Predef and PCO.

To have a more qualitative overview of the student activities, figure 4 shows the curriculum of each student through the activities made at 4 different times steps. For a given time step t and condition, a matrix slot represents the state of an activity (ordinate) for a particular student (abscissa). A slot is grey if a student has never explored the corresponding activity and it is purple if the student is doing this activity at time t. When a student has explored an activity, the slot is tinted green depending on the student's success rate (light green: low, dark green: high).

Across time, this figure allows to confirm the observation made previously which is that student working with ZCO and ZPDES are able to explore a larger set of activities than the ones working with PCO and Predef. We can also confirm that contextual choice does not affect the overall profile of activities proposed by ZPDES and Predef.

2.2 How does LP-based personalization impacts learning effectiness as compared to hand-designed curricula?

The learning effectiveness of each condition is evaluated through comparison between pre- and post-test results. The pre- and post-test (precisions in Sec. 2.2) are composed of 20 items scoring from 0 to 1 (max score is 20). The pre-test happens at the beginning of the first session, while the post-test happens at the end of the last session. The pre- and post-test are presented on the tablet on a dedicated interface (different from the ITS one). Each item of the test evaluates the student over knowledge and skills related to money manipulation, number composition, addition or subtraction (similar to the skills and knowledge trained in the ITS). Both tests include the same items organised in the same order but the items' wording have randomly selected values for each item and each student (with verification that no items in the post-test have the same values in their wording as the ones in the pre-test for one student).

The statistical procedure used in this section consists of three-way mixed ANOVA (algo x choice x pre/post) on the Math-tests score (pre- and post-measurement of student performance), with the pre/post factor as within-subject factor.

The algorithm factor includes the two conditions (ZPDES or Predef). And the choice factor include also two conditions (with or without choice). The p-value threshold is $\alpha = 0.05$. Pairwise comparisons are carried out with the Least Significant Difference (LSD) and Bonferroni procedure for corrected comparisons.

What is the impact of LP-based personalization without choice ? The main significant effect revealed an increase of the test score across time (pre/post factor effect, $[F(1,261) = 129.25,\ p-value = 0.000,\ \eta^2 = 0.331]$) which is boosted under the ZPDES condition compared to Predefined condition (algo x pre/post effect, $[F(1,261) = 40.076,\ p-value = 0.003,\ \eta^2 = 0.034]$). This effect combined with the examination of the marginal means (Predef: pre/post $mean = 6.83(sd : 0.353)$ / $8.36(sd : 0.396)$, ZPDES: pre/post $mean = 6.74(sd : 0.344)$ / $9.38(sd : 0.363)$) shows that children working with ZPDES algorithm learned more than the ones working with the Predefined sequence algorithm (visual support on Fig. 5).

Does the possibility to express choice boost learning ? Even more interestingly, the three-way interaction is significant (algo x choice x pre/post effect $[F(1,261) = 17.319,\ p-value = 0.049,\ \eta^2 = 0.015]$), the learning benefit from ZPDES condition is increased by the choice condition, whereas we observe the opposite for Predefined condition (Detrimental effect of choice). Pairwise comparisons indicate significant differences between PCO and ZCO for the post test score according to LSD procedure ($p-value = 0.014$), and only marginal differences according to Bonferroni procedure ($p-value = 0.08$).

2.3 How does LP-based personalization impact the emotional valence of learning experience and learner's motivation as compared to hand-designed curricula?

The emotional valence of the learning experience is assessed by an emotional scale. Thought this scale, the student can express how (s)he feels by moving a cursor from the best moment of his life to the worst one. The students answer this scale 3 times during each session (start, middle and end).

The statistical procedure used here consists of a two-way ANOVA (algo x choice) on the Emotional Scale Score (summative score of emotional valence of learning experience collected during each Kidlearn session, see Kidlearn-related learning experience in section 2.3).

The algorithm factor includes the two conditions (ZPDES or Predef). And the choice factor include also two conditions (with or without choice). The p-value threshold is $\alpha = 0.05$. Pairwise comparisons are carried out with the Least Significant Difference (LSD) and Bonferroni procedure for corrected comparisons.

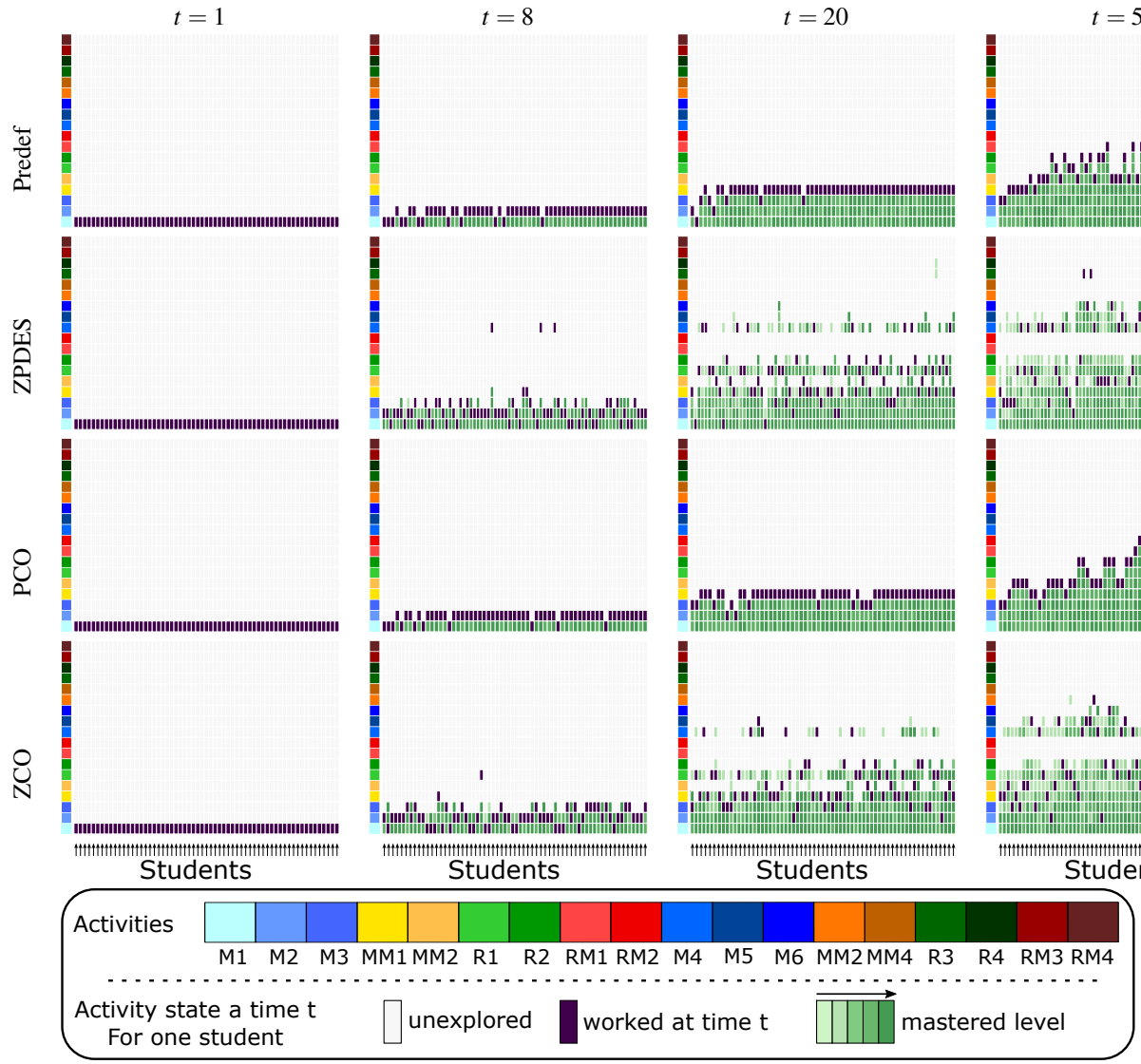


Figure 4. Students working with ZPDES and ZCO go through the graph of learning activities faster than students working with Predef and PCO. This way, they reach and achieve a larger set of activities.
There are 4 types of activity M, MM, R and RM with their related levels (M: 6, MM: 4, R: 4, RM: 4). They are ordered here in a colored band displaying a relative difficulty hierarchy to be able to facilitate the visualisation of the students' evolution across activities. Each cells represent the state of an activity for a student at time "t"; white sells for not explored, purple for activity done at time "t"; green for explored activity

The main significant effect revealed a difference for the Emotional Scale score between students who have choices and student without choice (Choice, $[F(1,261) = 12.060,\ p-value = 0.001,\ \eta^2 = 0.044]$). This effect combined with the examination of marginal means, (Choice: EmoScale $mean = 426.05(sd : 183.58)$, No Choice: EmoScale $mean = 338.86(sd : 226.96)$) shows that children working with the possibility to choose the object of the exercise feel better than the ones who does not have the possibility to choose, which suggests they are more satisfied of their leaning experience (visual support on Fig. 6).

Does the possibility to express choice boost motivation ? The motivation is evaluated through Vallerand's questionnaire[50,51]. It is based on Self-Determination Theory[52] and is commonly used to assess the elicitation of intrinsic and extrinsic motivation (e.g.[53]). It is composed of 21 items about the student's experience during the experiment sessions.

We also conducted a two-way ANOVA (algo x choice) on the Motivation score. The algorithm factor includes the two conditions (ZPDES or Predef). And the choice factor includes also two conditions (with or without choice). The p-value threshold is $\alpha = 0.05$. Pairwise comparisons are carried out with the Least Significant Difference (LSD) and Bonferroni

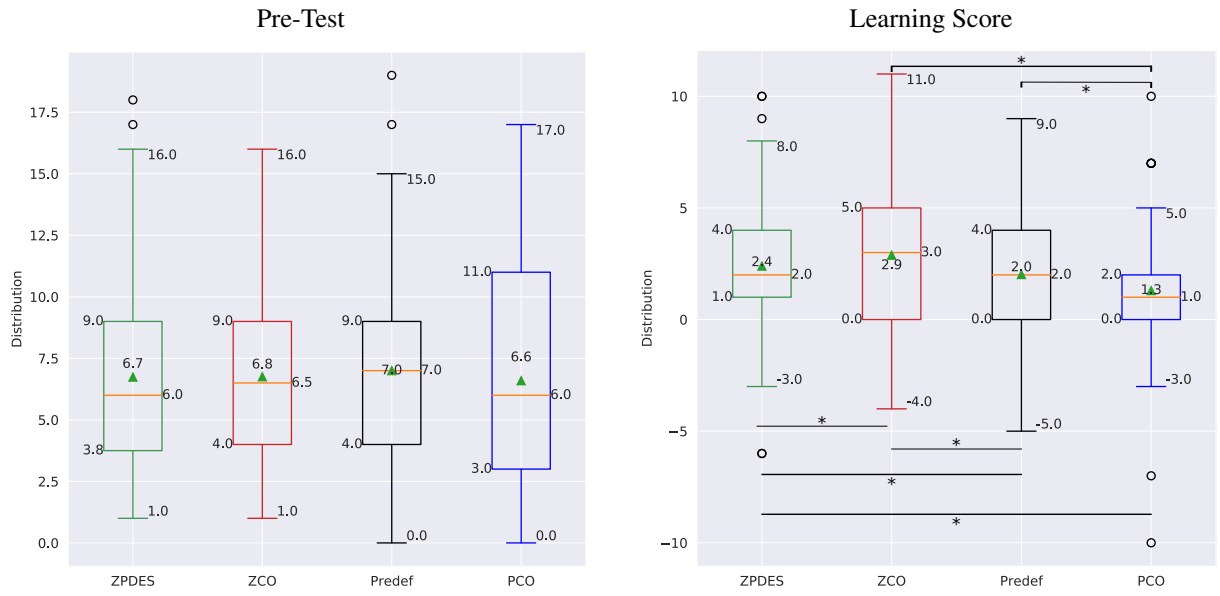


Figure 5. Boxplots presenting the Pre-test scores and the Learning score, i.e the difference between Post-test score and Pre-test score for the four conditions. The Pre-test scores are homogeneous among the populations, assuring a fair comparison. The Learning scores are ordered as follow ZCO > ZPDES > Predef > PCO, giving the order in term of learning efficiency between each condition.

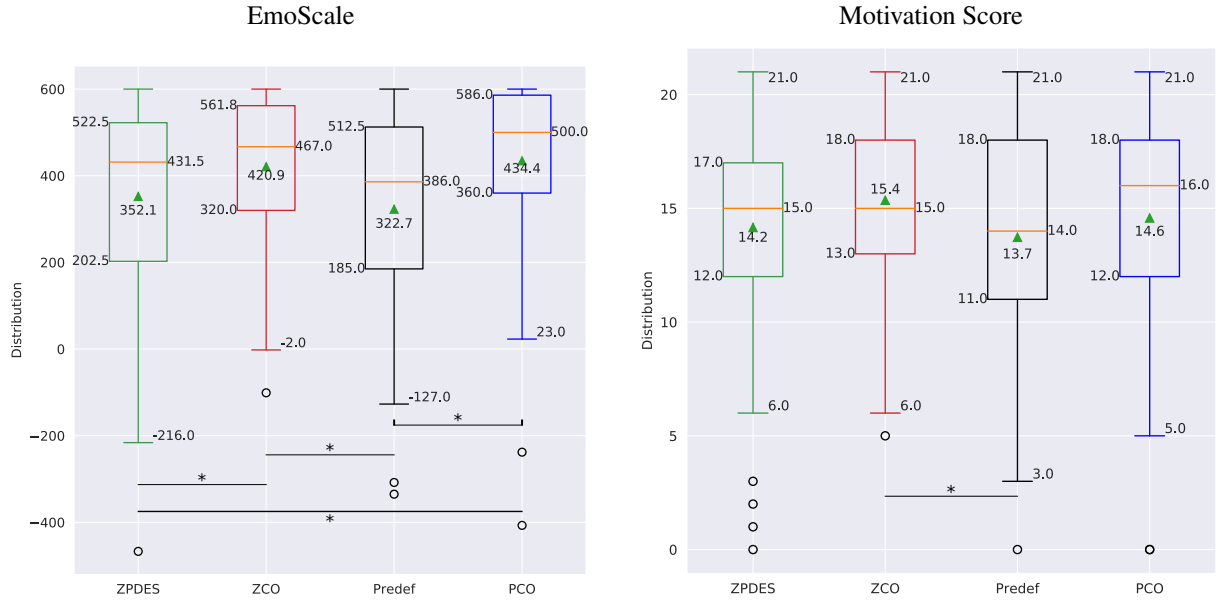


Figure 6. Boxplots presenting the Emotional Scale score on the left and the Motivation score on the right. Students working with ZCO and PCO show the highest EmoScale scores while students working with ZCO show the highest Motivation score, followed by PCO ad ZPDES and Predef present the lowest score.

procedure for corrected comparisons.

There is no significant effect but we can observe a tendency showing a difference between students who have choices and student without choice (Choice, $[F(1,261) = 3.449, p-value = 0.064, \eta^2 = 0.013]$). From this tendency, the examination of pairwise comparisons reveals a significant difference between ZCO and Predef according to LSD procedure ($p-value = 0.034$), but not from Bonferroni procedure ($p-value = 0.205$). This tendency combined to the examination of margin means (ZCO: MS $mean = 15.353(sd : 0.528)$, Predef: MS $mean = 13.726(sd : 0.553)$) seems to support that giving choice allows students to have a more motivating experience (visual support on Fig. 6).

This fits with the greater positive emotional experience elicited by the choice condition, and more particularly for ZCO condition.

2.4 Does a positive relation exist between LP-based personalization and subsequent learning performance and motivation ?

In order to establish a relationship between the learning effectiveness and the learning experience according to each experimental condition, correlations for each experimental condition (Predefined, PCO, ZPDES, and ZCO) were made between the following 3 measures (see Tab. 1) : 1) Learning score (difference between pre-and post-test) ; 2) Kidlearn progression (Final activity score defined in section 5.6.3) ; 3) Motivation score (see Sec. 5.6.3).

Importantly, the ZCO condition is the only condition where it is possible to observe positive relations between the learning score and the learning experience with the ITS in terms of both kidlearn progression and the motivation state (respectively $r = .21$ and $r = .27$). This observation correlates with the LP hypothesis that effective learning and intrinsic motivation are linked to activities in which learners progress and can exercise self-determination. Similarly, ZPDES condition induces a positive relationship between learning score and Kidlearn progression ($r = .32$). Taken together, these observed correlations strongly support the link between the progression in the Kidlearn app, enabled by this personalizing algorithm, and the actual learning progress.

In contrast, for the PCO condition, no correlation is significant (see table 1). This suggests that there is no link between the Kidlearn progression and the level of motivation elicited by the choice or the actual learning progress. As already mentioned, the children tended to be motivated by the choice opportunity but this motivation was not sufficient to actually progress while working with the predefined sequence.

Finally, for the Predef condition, no relation is observed between the learning score and the Kidlearn progression while the learning score under this condition is positively related to the student motivation ($r = .29$). As this condition induces the lowest learning score and the lowest motivation scores in children, this last correlation hints that learning outcome from Predef condition may be mainly related to the student's prior motivation, where the most motivated children do best, and the least motivated do worst, thus widening the differences in learning in this condition between the most motivated and the least motivated. This interpretation is corroborated by the much larger range of performances for these two variables in the predefined condition compared to the other three conditions.

Overall, this means that the ZCO is the best learning condition yielding actual learning progress associated to the learner's motivation. Additionally, only the LP-based conditions (ZPDES and ZCO) yield a reliable relationship between the progression across ITS based intervention and the real outcome in terms of learning benefit. In other words, the positive relation between LP and motivation is boosted when students can exercise their self-determination through choice in the ZCO condition. This solidifies the hypothesis that LP is correlated with/generates intrinsic motivation only when learner has the ability to choose, i.e. feels autonomous.

2.5 Is the impact of LP-based personalization modulated by individual characteristics of learners?

All the previous analyses have in addition been conducted with ANCOVA analyses where the covariable was related to individual characteristics. Particularly we investigated the mediating effect of school satisfaction, digital technology experience, gender and age (see Profil metrics in section 5.6.3).

No significant results have been observed revealing that the present results are robust to mediating effects related to the studied individual factors (see table 4 in appendix).

2.6 Synthesis

From the overall data, we can infer there is a double beneficial effect to the combination of ZPDES and choice on the learning and the motivational levels. This corroborates the results describing choice effect as a positive lever on motivation and performance[36–38]. However, our results show that the positive effect of choice, in terms of learning effectiveness, is algorithm-dependent. The choice is beneficial for ZPDES algorithm, whereas it is detrimental for predefined algorithm. This can be interpreted as, without a relevant teaching strategy, the choice will act as a distractor and students will focus more on the choice and less on the activity. In other words, allowing choice in inappropriate teaching strategies is deleterious for the students' learning, although they enjoy to make choices.

		Kidlearn-related learning experience	
Learning score (Pre/post difference)		Kidlearn progression (Final activity score)	Motivation score
Predef	R value	.001	.**29**$^{+}$
	P value	*ns*.	.02
PCO	R value	-.006	.11
	P value	*ns*.	*ns*.
ZPDES	R value	.**32**$^{+}$	.07
	P value	.005	*ns*.
ZCO	R value	.**21**$^{+}$	.**27**$^{+}$
	P value	.07	.02

Table 1. Bravais-Pearson inter-correlation between Learning score (pre/post difference) and the Kidlearn experience scores with the ITS application (Kidlearn progression and the motivation score). Notes. ns. = non significant. According to Fisher's transformation procedure (with the limit values for Z at 1.96), r values comparisons revealed no significant difference across conditions

3 Discussion

On a large sample of students, our results clearly indicate that personalization of the learning path via an algorithm that estimates the proximal learning zone by maximizing LPs is more effective in terms of learning outcomes than "linear design" strategies that only adapt the pace and number of exercises across the linear curriculum. This result is consistent with a systematic review on ITS ([46]), indicating that personalization is more effective than one-size-fits-all instructional design and that ITS are more effective than traditional whole-class instructional methods.

Specifically, for the first time, we report an extensive study showing that personalizing the pathway according to the student's LP improves learning performance while producing a positive and motivating learning experience, regardless of several learner characteristics such as gender, experiences with technology, past experiences with the activity being trained, or the student's experiences and perceptions of school. Taken together, this empirically supports the robustness of LP-based personalization to diverse student characteristics (known to be critical to learning). Also, in line with our hypotheses, we show for the first time also the added value for learning outcomes of associating the LP-based individualization of the learning path with a playful feature allowing self-determined decisions, yielding a synergy of intrinsic motivations elicited by both the LP (as assumed in the LP hypothesis,[24,26,28,30]) and by "gamification" strategies (e.g.,[54,55]).

It is noteworthy that this positive synergy on the learning outcome is associated with a positive and motivated learning experience. As a result, in the ZCO condition, instructional effectiveness (pre-/post difference) is also positively correlated with motivation scores.

Conversely, we show a deleterious effect of the association of a playful feature with a linear learning pathway in terms of real learning outcome contrasting with a positive and motivated learning experience for the students. As a result, the correlation between learning score (pedagogical effectiveness) and intrinsic motivation was not significant. This result is particularly insightful because it highlights that a positive and motivated learning experience via a "gamification" strategy that elicits intrinsic motivations, is not sufficient to improve learning outcome. In other words, in that case the attention-grabbing power of games can lead to a distraction from the pedagogical objectives of the activity.

This detrimental effect of "gamification" on learning performance under the PCO condition mirrors findings in children about the motivational conflict between immediate and delayed rewards (also called want-should conflicts ([56–58]). In our case of PCO condition, the choice of object for each exercise can be seen as an immediate reward (without learning gain expected) while the learning progression into the kidlearn are delayed rewards not very attractive due to their small magnitude related to the "one-size-fits-all" design of this condition. Overall, this reversal effect of "gamification" for linear learning path invites to be cautious when using "gamification" strategies for teaching purposes. Today, one of the great challenges of modern education is that of capturing the attention of students and creating engagement for learning tasks. In light of our results, using "gamification" strategies to enhance motivation and learning is effective only if ITS features actually fosters learners' learning progress as provided by our LP-personalization. Such a result is consistent with the self-determination theory applied to education stressing intrinsically motivated learning for really meeting learners' autonomy and competence needs ([22,59]).

Finally, a very salient result to highlight are the correlations observed between the learning score (pre-/post difference, i.e. progress) and the learning progression within the Kidlearn ITS. Positive relationships are observed for the two conditions with LP-based personalization, but not for the two conditions with linear pathways. Hence, the learning progress observed post-intervention is really linked to the learning progress obtained through LP-based personalization. In contrast, such an

assertion is not possible for the two conditions with linear pathways since the correlations are not significant. Indeed, the linear pathway with gamification seems to have made children less focused on the learning task as explained above and the linear path alone seems to be the less motivating, thus, the learning progression in this conditions seems to reflect a combination of prior level with very little effective learning from activities and a little bit of learning from test-retest learning effect ([60]). Consequently, it can be argued that tools for visualizing learning paths with ZPDES or ZCO, as well as giving feedback to the student, or an instructional monitoring interface for teachers, has the potential to give reliable hints on the student's learning progress .

4 Conclusion

The present field study assessing our ITS approach of personalization driven by learning progress provides conclusive results in terms of both pedagogical effectiveness (progress observed pre- and post-intervention) and efficiency (learning experience and motivation elicited post-intervention). Indeed, LP-based personalization (ZPDES driven) provides better learning outcomes and a better learning experience than a linear-path sequence. Furthermore, we observe a synergic effect between LP-based curricula and the ability to express choice (allowing to express self-determination), in accordance with the LP-model. On the contrary, allowing choices, as a form of gamification, can have a deleterious effect and act as a distractor when combined with linear-path curricula.

Other results in field studies have shown the LP approach to be effective for students with specific learning needs (i.e., Autism and intellectual deficiency,[61]) and applicable to different domains (health education,[62]) showing the generality and promising perspectives of the approach. Future work could also evaluate the use of this approach in the field of cognitive training to improve the number of responders to training in various samples (age, neurodiversity).

5 Materials and Methods

Several systematic reviews or studies on ITS[1,3,45,48,49] efficacy pinpointed methodological limitations of this new empirical field (no control group, no initial group equivalence, no pre- and post-intervention measurements, etc[47]) and the great variability of the ITS designs or of their use making it difficult to identify which of the ITS features and/or which of the conditions of learning context of their use [2] are critical for successful personalized learning. So in this section, we describe both the ITS features as well as the experimental protocol. As the ITS system presented here was conceived in the context of a project called 'KidLearn', we also refer to this ITS as the KidLearn system.

The purpose of an ITS is to enable a learner to acquire knowledge and skills related to a specific domain. Modelling such domain is a difficult problem that has been the subject of numerous research[63–65]. Difficulties also arised during the research around the use of Q-matrix for the use of Multi-Armed Bandit (MAB) for ITS[34]. Even if a lot of research has been done to create tools such as Cognitive Tutor Authoring Tools (CTAT)[66] to help experts create Q-matrix[67], their use in the conception of the domain model can lead to practical difficulties such as human errors, misspecifications[68] and heavy time consumption for the pedagogical expert. The following Activity Space formalism is defined to address these issues.

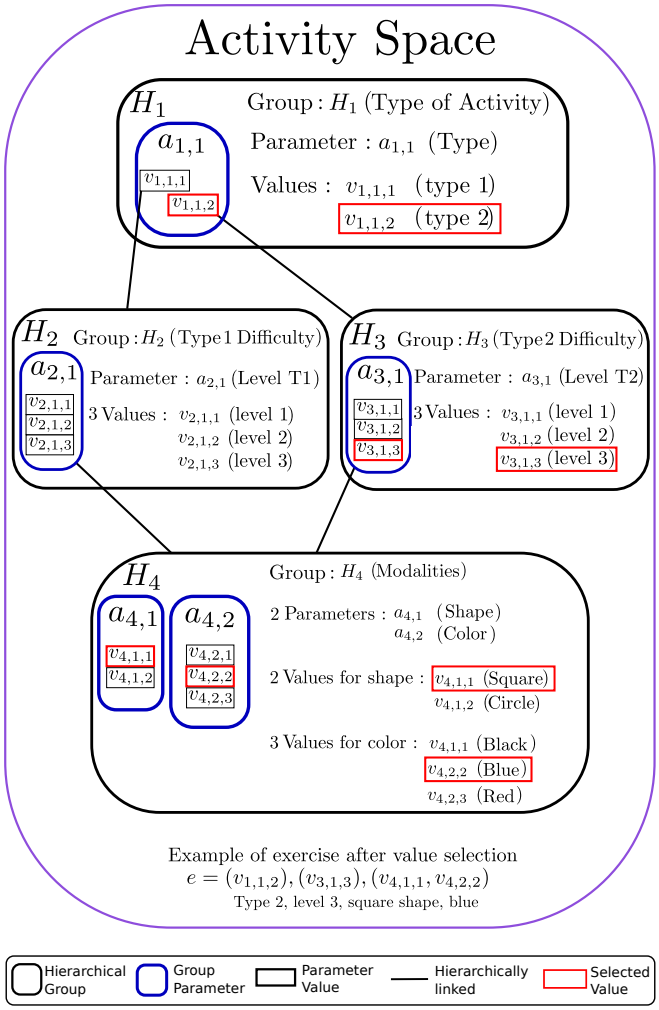
5.1 Activity Space definition

A pedagogical Activity Space is considered to be a set of activities that a learner can practice to acquire skills or knowledge components. An activity or exercise is characterized by multiple parameters a_i (difficulty, shape, type, ...) which can take different values v_j. For example, to work on mathematical skills, an exercise may have a type that works the addition and another type that works subtraction. "Addition" and "subtraction" are then two possible values for the parameter "type of exercise".

These parameters and their respective values define all the possible activities that can be instantiated inside the activity space. Depending on their nature and meaning, these parameters can be organized in different groups. Such group of parameters is noted as $H_x = a_1, \ldots, a_{n_x}$. In addition, these parameter groups can be structured hierarchically, since some parameters depend on others to be used in an activity.

Different types of exercises can require different skills, so the first group of pedagogical parameters (which will be the first level in the hierarchy) determines which type of exercise is selected. Several difficulty levels exist for each type of exercise, so different groups of parameters will determine which difficulty is chosen depending on the type of exercise (second level in the hierarchy). In this case, when one exercise type is selected, the parameter groups that determine the difficulty for the other types are not involved in the parametrization of the activity. Thus, not all parameter groups are necessarily used to define all the activities in the activity space. Therefore, an Activity Space is defined as a set of n_H hierarchical groups of parameters, $A = H_1, \ldots, H_{n_H}$.

[2] for instance, ITS can be used alone or mixed with a specific teacher-based instructional setting



To compute an exercise, the values are first selected in the primary group H_1. Here 2 values are possible for 1 parameter. $v_{1,1,2}$ is selected thus $u_{1,1} = 2$ and the first group instantiation gives $h_1 = v_{1,1,2}$, meaning the exercise type is 2. This value is hierarchically linked to the group H_3. (H_2 is not used in this case).

Then, values in the group H_3 are selected. Here, $v_{3,1,3}$ is selected so $h_3 = v_{3,1,3}$, i.e. a difficulty level of 3. All levels are linked to the group of modalities H_4. Next, values are selected in the group H_4. For example $v_{4,1,1}$ and $v_{4,2,2}$ leading to $h_4 = v_{4,1,1}, v_{4,2,2}$ meaning a square shape and a blue colour.

So the final exercise resulting from the values selection is:
$e = h_1, h_3, h_4 \leftrightarrow e = (v_{1,1,2}), (v_{3,1,3}), (v_{4,1,1}, v_{4,2,2})$

This means the exercise is of Type 2, level 3 with square shape and blue modalities. The Activity Space groups all the activities that can be generated in this way.

Figure 7. Illustration of an Activity Space with 4 groups of parameters, and a selection of values which lead to an example of an activity. A group is noted H_x, a parameter $a_{x,i}$ and a value $v_{x,i,j}$.

An activity/exercise e is characterized as a particular combination of parameter values inside an activity space where values were selected for each hierarchical group of parameters involved. All the parameters needed to define an activity are instantiated to produce a unique combination of parameter values. The index u_i corresponds a selected value v_{u_i}, for a parameter a_i, used to generate an activity. To simplify the notation, u_i is noted as a given parameter selected value to generate an exercise to differentiate it with v_j which defines any values of a parameter. For a group H_x with m parameters, the selection of each parameter value produce a combination leading to a singular instantiation of this group $h_x = u_1, \ldots, u_m$. After the selection process, a certain number of groups was instantiated, each producing an activity $e = h_1, \ldots, h_{n_e}$, which groups all parameter values that were selected to produce a unique combination. An activity space groups all possible distinct combinations of parameter values that can define an activity in this space. An illustration of a simple example of an Activity Space with the instantiation of an exercise is shown in figure 7.

How can this activity space be managed to propose relevant and personalized activities and offer a motivating and enriching experience to the learners ? Several methods are proposed below to answer this question.

5.2 ZPDES : a combinaison of Multi-Armed Bandit and Intrinsic Motivation theories to manage Teaching Sequences

To address the challenge of managing activities in an Intelligent Tutoring System, the ZPDES (Zone of Proximal Development and Empirical Success) has been proposed as an implementation of the HMABITS architecture[69]. It relies on state-of-the-art Multi-Armed Bandit techniques (MAB)[32,33] and exploit tthe empirical estimation of learning progress[70] to manage the Activity Space.

To use a casino analogy, multi-armed bandits describe the problem of finding the slot machine that provides the maximum

reward, initially unknown, in a set of many different machines. To find the best machine it is needed to spend money exploring each one before being able to always bet on the best one. This boils down to what is called the "exploration/exploitation" trade-off in machine learning and learning processes generally. Here, these approaches are adapted to ITS where the gambler is replaced by the activity manager, the choice of machine is replaced by a choice of activity parameter values, and the reward is replaced by the student learning progress. It is assumed that activities which are currently estimated to provide a good learning progress must be selected more often as described in section.

A particularity here is the reward (learning progress) which is non-stationary. This requires specific mechanisms to track its evolution. Indeed, a given activity will stop providing a reward, or learning progress, after the student reaches a certain mastery level of the skill or of the activity. Also, it cannot be assumed that the rewards are independent and identically distributed as different students will have different preferences, sensibilities or human factors. They may be distracted or make mistakes when using the system which can create spurious effects. Thus, the framework introduced here rely on a variant of the EXP4 algorithm, proposed initially by[32], which considers a set of experts[3] and make a choice based on the proposals of each expert. In case presented here, the experts are a set of variables that track how much reward each activity is providing[72]. These bandit experts are used to evaluate the quality of each activity parameter value during the learner's working session.

Due to the combinatorial explosion of parameter values, only one MAB is not used for each possible combination of parameters values in the activity space but a set of simultaneous MAB is used for each group of parameters. The first alternative of considering a given arm for each activity would increase the number of arms. That would increase the number of parameters and the number of trials required to estimate learning progress and thus the learning time. Also, the approach presented here allows the algorithm to identify which features benefit some students more than others.

For example, to learn a particular skill, the same information may be presented in a written text, a video, a game, an audio track or another format. The knowledge the learner must acquire is the same in each case, but the format of the information differs and individual learners may be more receptive to a particular format.

A case can be imagined where a student works to learn mathematics; different activities are presented to him in a written format, and he almost never answers correctly. But when activities are presented to him in an audio format, he begins to succeed and progress. In this case, the problem is not about the mathematics skills he could learn, but rather his skills in reading. As another example, if an audio format is presented to a student with a hearing impairment, he will not perform and progress as well as with a written format. In light of this, the introduced method evaluates the relevance of and gives meaning to each feature and detects weaknesses and preferences of each student. The propositions it makes are more customized than the ones from an approach where particular combinations would be evaluated, but where features are not taken into account.

Each simultaneous MAB, used to sample each group of parameter, uses a bandit algorithm derived from EXP4[72]. The following process is described in Alg. 1.

Algorithm 1 Procedure to stochastically sample group parameter values according to their quality evaluation.

Require: Group H_x of m parameters a_i with their n_i values v_j
Require: Set W_x of m experts w_i for each parameter
Require: γ rate of exploration
Require: distribution for parameter exploration ξ_u

procedure SAMPLEVALUES(H_x, W_x)
  for $i = 1 \dots m$ **do**
    $\tilde{w}_i \leftarrow \frac{w_i}{\sum_{j=0}^{n_i} w_i(v_j)}$
    $p_i \leftarrow \tilde{w}_i(1-\gamma) + \gamma\xi_u$ (Eq. 1)
    $u_i \leftarrow$ value sampled from a_i proportionally to p_i
  end for
  $h_x \leftarrow \{u_1, \dots, u_{n_x}\}$
  return h_x
end procedure

For each parameter a_i inside a group, the quality of its values is evaluated by a bandit expert w_i. An expert track the reward provided by each value v_j on the last several sampling to compute its quality noted $w_i(v_j)$. At any given time, the value to use for each parameter is sampled according to the probabilities given by:

$$p_i = \tilde{w}_i(1-\gamma) + \gamma\xi_u \tag{1}$$

[3]The general term "expert"[71] is used to refer to strategies used in algorithms for"prediction with expert advice", "by combining the predictions of several prediction strategies".

where $\tilde{w_i}$ are the normalized w_i values to ensure a correct probability distribution, ξ_u is a uniform distribution that ensures sufficient parameter exploration and γ is the exploration rate, tuned to make the exploration wide or narrow. This sampling methodology leads to stochastically select a value, proportionally based on its quality and γ. For low values of γ, the parameter value is chosen mostly based on its quality, whereas for high values of γ, low quality parameter values have a higher probability of being picked, which means a high exploration rate. The set of experts correlated to H_x is noted $W_x = w_1, \ldots, w_{n_x}$. From now on, a Stochastic Activity Space A^S is considered to be a set of tuples (H_x, W_x).

To generate an activity, this process is done recursively on the hierarchical groups that are involved in the activity generation, in accordance with the hierarchical dependencies between the groups of parameters. As describe in Alg. 2, it starts by the instantiation of the primary group of parameter H_1 and is followed by the instantiation of the groups that are iteratively selected according to their dependencies. This leads to a stochastic draw of activity, resulting from the combination of each parameter value sampled depending on the evaluation of their quality by each expert. An abstract illustration of an activity generation is presented in figure 8.

Algorithm 2 Activity generation procedure based on an Activity Space and Hierarchical Multi Armed-Bandit mechanisms.

Require: A Stochastic Activity Space A^S, set of tuples (H_x, W_x)
procedure GENACTIVITY(A^S)
    {Initialize}
    Instantiate primary group $h_1 \leftarrow sampleValues(H_1, W_1)$
    $i \leftarrow 1$
    {Recursive sample}
    while h_i require to instantiate a group H_x **do**
        $i \leftarrow x$
        $h_i \leftarrow sampleValues(H_x, W_x)$
    end while
end procedure
return $e = h_1, ..., h_i$
[h]

Once an activity is generated, this activity is proposed to a learner to work on and answer to. After answering, the algorithm retrieves his answer. Each time an exercise is given and answered, the expert of each parameter value u_i used in the activity is updated:

$$w_i(u_i) \leftarrow \beta w_i(u_i) + \eta r \tag{2}$$

where r is a reward that measures the benefit the activity gives to the learner in terms of progress. The variables β and η define the tracking dynamics of this estimation, which is the compromise between the old rewards and the new ones brought by the last activity. This mechanism allows the experts to assess and update the quality of each parameter value, used over time, based on the student learning.

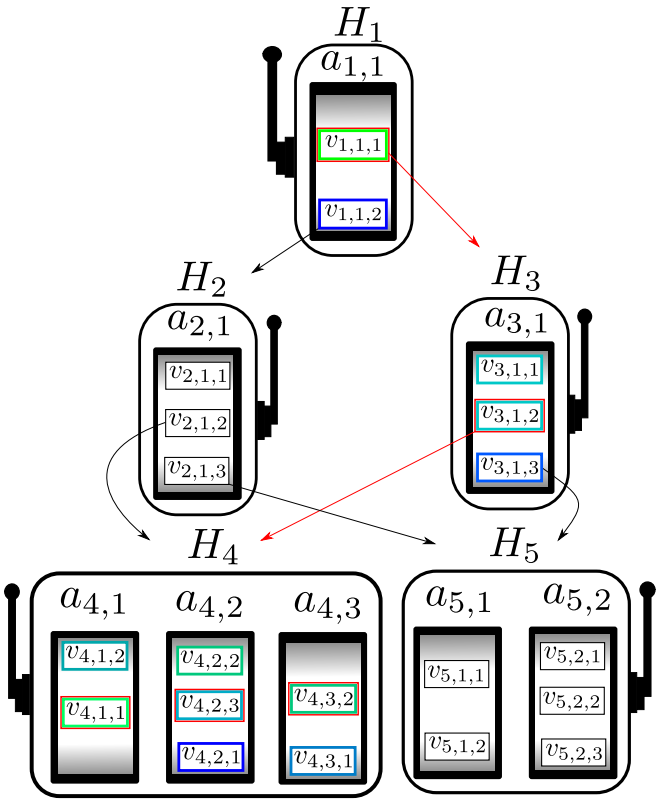
As discussed before, focusing on activities that are providing more learning progress can act as a strong motivational cue[21]. Equation 3 describes the reward computation which is based on estimating how the success rate on each parameter group is improving :

$$r_x = \sum_{t=T-d/2}^{t} \frac{C_t}{d/2} - \sum_{t=T-d}^{T-d/2} \frac{C_t}{d-d/2} \tag{3}$$

where $C_t = 1$ if the activity at time t was solved correctly. At the time T, the equation compares the success of the last $d/2$ samples with the $d/2$ previous samples, providing an empirical measure of the time evolution of the success rate.

This reward allows to compute a measure of the quality of each activity parameter value, measuring how much progress it provided in a recent time window. Both extreme cases, when an activity is already mastered or when it is impossible to solve, will have a reward of zero. Moreover, parameter values providing a faster progress are assumed to be better than others. The algorithm to compute the reward is presented in Alg. 3.

A pure selection, based solely on the previous considerations, would explore all possible activities that could be generated in the activity space from the start of the work process. This would have two drawbacks. First, the type and difficulty of the exercises proposed could change too often and reduce the learners' motivation and engagement. Second, it might not be possible to explore all activity parameters to estimate the learning progress they are providing. To ensure that learners



The primary group H_1 has one parameter $a_{1,1}$, for this parameter, the first values is evaluated to have a higher quality than the second one: $w_{1,1,1} \geq w_{1,1,2}$. There are then more chances for the first value to be sampled. Here, the result of the stochastic sample is that $v_{1,1,1}$ is drawn.
$v_{1,1,1}$ is linked hierarchically to H_3, which is the next to be instantiated. The two first values have the same medium quality (same chance to be drawn), and the last one has a low quality (less chance of being drawn). $v_{3,1,2}$ is drawn, which has a dependency with the group H4.

H4 is then instantiated and has three parameters. The first parameter $a_{4,1}$ has its first value evaluated to be more interesting than its second one. The quality for the second parameter is ordered as $w_{4,2,2} \geq w_{4,2,3} \geq w_{4,2,1}$ and for the third parameter $w_{4,3,2} \geq w_{4,3,1}$.

The three parameters are sampled simultaneously, and $v_{4,1,1}$, $v_{4,2,3}$ and $v_{4,3,2}$ are drawn. Even though $w_{4,2,2} \geq w_{4,2,3}$, $v_{4,2,3}$ had a chance to be drawn, following the process of exploration. The result of the activity generation is:
$e = (v_{1,1,1}), (v_{3,1,2}), (v_{4,1,1}, v_{4,2,3}, v_{4,3,2})$.

Figure 8. Hierarchical Multi-Armed Bandit with 5 groups of parameters and a selection, by stochastic draw, of an example of activity. A group is noted H_x, a parameter $a_{x,i}$ and a value $v_{x,i,j}$.

Algorithm 3 ZPDES reward computing procedure

Require: Activity e
Require: Student answer C
Require: parameter d
procedure COMPUTEREWARD(e, C)
  for h_x in e **do**
    $r_x = \sum_{t=T-d/2}^{t} \frac{C_t}{d/2} - \sum_{t=T-d}^{T-d/2} \frac{C_t}{d-d/2}$ (Eq. 3)
  end for
  $r \leftarrow r_1, \ldots, r_{n_e}$
  return r
end procedure

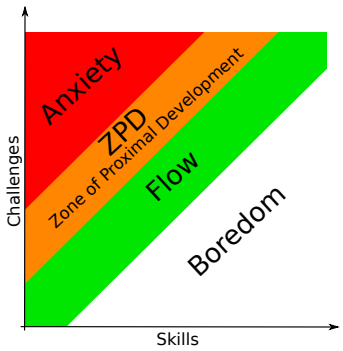


Figure 9. The concept of Zones of Proximal Flow[73] presents the idea of the Zone of Proximal Development being located in between regions of Flow and anxiety.

remain in challenging but possible to achieve areas and to be able to assess the quality of each parameter, a mechanism to limit exploration is introduced. Inspired by the Zone of Proximal Development theory[15] and the concept of Flow[16], a pedagogical expert has the possibility to specify rules that define an evolving set of possible/activated activities, judged relevant for the student. These activities keep the student in the zone of Flow or in the Zone of Proximal Develpment (ZPD) based on his successive results (see Fig. 9). The goal is to propose activities that are neither too easy nor too difficult, without having to try all possible activities. The different possible activities proposed by the algorithm are then the active ones which are inside the ZPD. The use of the ZPD offers three advantages: it helps to improve motivation as discussed before, it further reduces the need of quantitative metrics for the educational design expert and it provides a more predictive choice of activities.

The implementation of these principles is applied to the algorithm by the definition of rules that guide the bandits experts and restrict the exploration of the activity parameters. Theses rules define activation/deactivation mechanisms which allow the algorithm to activate and deactivate parameters values, depending on the evaluation of their relevance and the quality of the students learning process. As a consequence, the active parameters values generate a subset of all possible activities inside the activity space. The ZPD is defined here as a particular subset of active activities with its corresponding parameter values.

Following this principle, ordered relations between activity parameter values can be defined, leading to a "graph" governing the activity space which are combined with the set of rules that define and manage the ZPD using activation/deactivation mechanisms. Rules and ordered relations are not always defined for each parameter: there is a distinction between subsets of activity parameters that have a clear difficulty progression, and subsets that don't. For the example used from Sec 5.1, the difficulty levels have a clear ordering while the modalities don't. In practice, the management of the ZPD proceeds as follows. For activity parameters with no difficulty level relations between their values, a free exploration is allowed and so all of their values are always active. While for parameters that have a clear progression in difficulty, the values will be activated and deactivated depending on the success rate over all active values.

The following mechanism is proposed to generally manage the ZPD. When the recent learner success rate over all active parameter values $\delta_{i,ZPD}$ reaches a value λ_{ZPD}, the ZPD is expanded to explore another parameter value $v_{i,j}$ by initializing its expert as : $w_i(v_{i,j}) = \min w_i(v^{ZPD})$. When the recent success rate for a particular value $\delta_{v_{i,j}}$ is higher than a threshold λ_d, this activity can be deactivated and removed from the active list of values. These two threshold allow to configure the general exploration behaviour of the algorithm inside the activity space and is illustrated in figure 1.

The main intuition of this process is that when there are some activities whose difficulty grows, the ZPD will have to grow at the same rate. When activities do not have a clear order of difficulty, or when the order might change from person to person, then it is necessary to allow wider exploration of the activities to accommodate individual differences.

Another kind of mechanism is added to allow a more precise and specific parametrization of the ZPD. Indeed, the algorithm needs to be able to activate and deactivate values when the conditions of exploration for an activity parameter depends on another set of parameters. If the value $v_{g,i,j}$ of parameter a_i of group H_g requires a certain mastery level of value $v_{x,y,z}$, a threshold $\lambda_{v_{x,y,z}}$ is defined corresponding to the success rate a learner must reach with activities using $v_{x,y,z}$ to activate $v_{g,i,j}$. The requirements can be multiple, meaning a values activation can depend on multiple other values, parameters or group to be activated.

For example, the difficulty level for a particular type of exercise can require only the previous level to be mastered, or it can require various previous levels, or it can even require other types of exercises in different levels to be mastered. In a mathematics analogy, if a student works on a simple subtraction activity, he needs to master simple addition activities to be able to succeed. And if he works on hard subtractions with basic decimal number, he needs to master hard addition and basic decimal numbers first to succeed.

But this mechanism can lead to blockages in the exploration. If a type A of exercise is easy during 3 levels for a student, leading to a 100% success rate, the quality of this type will be very low. The algorithm will then select other types of excercises more often. But if the level 2 of type B needs a higher level of type A to be mastered, the algorithm will continue to propose more type B without being able to activate the level 2 until the required level type A is mastered.

To address this issue, a quality upgrading mechanism is added for values that are required. If ZPDES tries to activate a value parameter but is unable to do so due to a requirement, the qualities of required parameter values are increased. This way, ZPDES will exploit values needed to expanded the graph as a priority.

Basically, the first mechanism introduced to manage the ZPD is a simplification of the mechanism presented above. It is integrated to reduce the information needed to define ZPD rules and to allow a freer exploration of the activity space by the algorithm.

Algorithm 4 ZPDES algorithm. It manages pedagogical curricula based on an ActivitySpace, a multi-armed bandit algorithm (genActivity procedure), and a set of rules to extend and/or shrink the ZPD where the student evolves.

Require: A Stochastic Activity Space A^S
Require: R^{ZPD} rules
Initialize bandit experts uniformly according to R^{ZPD}.
while *learning* **do**
  Generate activity $e \leftarrow genActivity(A^S)$ (Alg. 2)
  Get learner answer C
  Compute reward $r \leftarrow computeReward(e, C)$ (Alg. 3)
  Update greedy expert
  for (h_x, r_x) in (e, r) **do**
    for u_i in h_x **do**
      $w_i(u_i) \leftarrow \beta w_i(u_i) + \eta r_x$
    end for
    Update ZPD: activate/deactivate w_i based on R^{ZPD}
  end for
end while

The final ZPDES algorithm is presented in Alg. 4. One of the main advantages of these principles is the consideration of an empirical estimation of the learning progress. It has been proposed in artificial curiosity and intrinsic motivation systems[70]. Instead of relying on a precise model of the learning system, with all limitations in terms of parameter identification and computational complexity, it is possible to create surrogate functions of the learning progress. These estimators are simple, robust, and, even if not optimal, more flexible and adapt better to model errors and situations where the model assumptions are violated.

Added value of the LP based approach Presenting the best activities to a learner at a given time to stimulate his learning as well as motivation is a crucial issue in ITS design. Therefore, the evaluation of the student knowledge level, i.e the "student model" (and subsequent adaptations), needs to be accurate over time to provide the best match between the learning activity and the learner's zone of proximal learning ([74]).

The general principle of the LP-based approach is to propose to each learner the activities that maximize his or her progress within the ITS activities. Such an adaptation is dynamic and depends on the learner's performance.

The activities are structured as a graph into an activity space based on expert knowledge (i.e, the "domain model"). The learning paths are then personalized in two ways; first, by exploring and testing continuously various activities inside the activity space in order to assess their didactic potential for the learner's progress in real time; second, by exploiting and mainly proposing the activities identified as being the most effective for him/her based on the previous assessment.

The simplicity of the activity space on which the approach is based is a first asset. Indeed, it does not rely on any multidimensional student or domain models, and then only the learner's information about the estimated learning progression for each activity is required. A second asset of LP approach is to leverage an efficient and simple optimization method consisting of prioritize activities' parameters identified as yielding significant learning outcomes, i.e. Zone of Promixal Development (ZPD). For that, thanks to our multi-armed bandit algorithm, at each step of the optimization process, one arm is chosen and the resulting payoff is estimated, the objective being to discover dynamically the best arm for each student.

So, various and heterogeneous paths are possible across students as expected for taking into account specific learner's need at a specific point in time. So, the LP approach empowers the ITS adaptivity for diagnosing learner's ZPD and for making appropriate adjustments to the specific learners' needs. Taken together, these two main assets enable to quickly design

successful ITS (albeit substantial efforts must be done to parameterize the activity graph on which ZPDES runs) and are supportive to hybrid system combining machine learning and rule-based approaches[75].

5.3 Predefined Sequence

To be able to evaluate ZPDES, a baseline has been built as the form of a Predefined Sequence (Predef). This Predefined Sequence is a simple algorithm inspired by mastery learning strategy[76] and instructional design whose reliability has been validated through several user studies[77] and it does not use any machine learning technology. It consists of a sequences of predefined activities organised by difficulty. When a student is working an activity, he needs to have 3 success out of 4 exercises to pass to the next activity, or else he stays on the same activity. The sequence of activities used in this experiment as been designed by an expert in teaching of mathematics for primary school. The sequence is composed of 27 activities we can divide into 8 groups. Each group corresponds to a type with or without decimal. (M, R, MM and RM with integers only, then M, R, MM and RM with decimals).

5.4 Choice

The ability to make one's own decisions, i.e. the ability to make choices, is part of the learning-progress theoretical framework[20,24,78] on which the development of the ZPDES algorithm was scaffolding. Choice expression was also shown to have a positive motivational impact and an efficient vector of performance[36–38]. As ZPDES and Predef do not actually enable students to express choices (which are performed by the machine learning algorithm), two conditions are introduced to reintroduce this ability and study its impact. PCO and ZCO do not change the way ZPDES and Predef control the evolution of parameters of learning activities, but they introduce contextual choice on the objects used to instantiate visually the learning activities the students train as presented in figure 10. The aim is to increase intrinsic motivation by adding a preference depending on the student personality and thus introducing an emotional and motivational valence on the object[79].

Figure 10. Object choice interface. The student choose the object(s) he wants to train with by typing on it. The bubble indicate instruction "Choose what you want".)

As explained before, studying the impact of choice as a motivational and learning tool, and decoupling it with the actual pedagogical content (e.g. difficulty of exercises) are interesting. ZCO is an experimental condition where the student has choice over a contextual parameter: the objects presented on the screen. In the money game scenario (see below Sec. 5.5), students compose sums corresponding to the item prices or the change to be given when a customer purchases an item. The choice given to the student is between two different objects, but the activity parameterization is the same. The activity is still selected by a ZPDES algorithm and the choice has no impact on the ZPDES operation. The interface used to implement this experimental condition is shown in figure 10. Only the type of exercise and the objects are presented to simplify the interface and reduce the perturbation of the student to a minimum. The position of the choice icon on the screen is determined randomly to avoid presentation bias.

5.5 Kidlearn activities scenario

The teaching scenario used here is about the use of money to teach children how to decompose numbers, typically targeting 7-8 year old students. It corresponds to a set of mathematical skills and learning scenario that are part of the official learning curriculum of French primary schools for children of this age range. This scenario was chosen for its simplicity, while remaining rich enough to offer different learning/teaching trajectories to impact individual students differently. The entire conception of exercises, ranging from their parameterization to the visual interface, was conceived in collaboration with a specialist of didactics of mathematics and participatory design of primary school teachers.

Furthermore, combining number and money manipulation is a way to instantiate abstract knowledge into a practical, useful real-world scenario. This scenario is instantiated in a browser environment.

The application proposes exercises to students in the form of money games (see Figure 11). For each exercise type, one object is presented with a given tagged price, and the learner has to choose which combination of bank notes, coins or abstract tokens need to be taken from the wallet to buy the object, with various constraints depending on the exercise parameters.

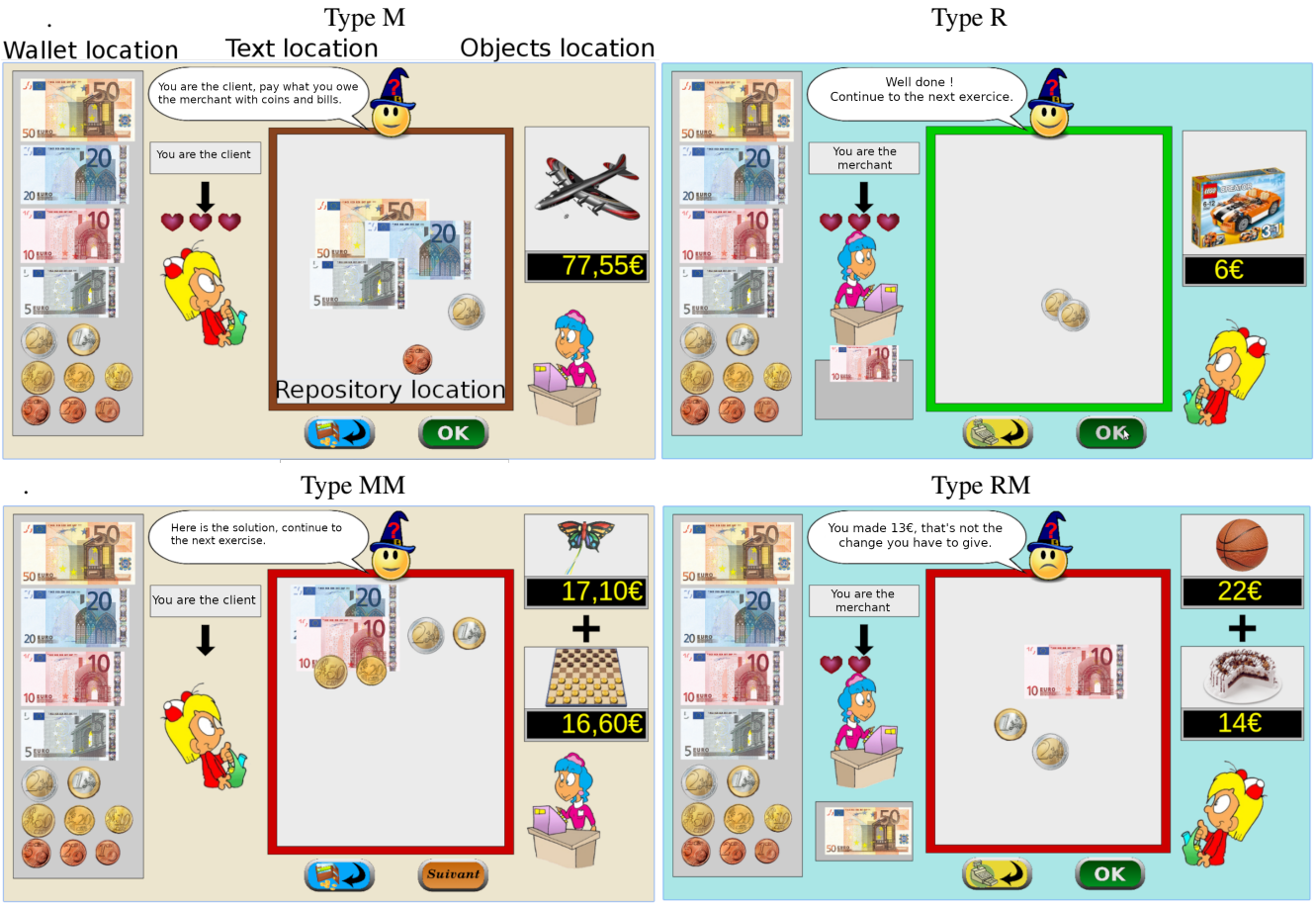


Figure 11. Four principal regions are defined in the graphic interface. The first is the wallet location, where users can pick and drag the money items and drop them on the repository location to compose the correct price. The object and the price are present in the object location. Four different types of exercises exist: M : customer/one object, R : merchant/one object, MM : customer/two objects, RM : merchant/two objects.

The various activities are parameterized using a specific graph summarized in figure 12 with an example of ZPDES activity sampling. There are 5 parameters organized hierarchically. First, the **Exercise Type** is chosen: the student can be the costumer or the merchant and buy or give change with one or two objects, which leads to four different possibilities. For each type of exercise, the difficulty is chosen based on the difficulty **Level** of decomposing a number. A number can be easy to decompose if there is a direct relation with a real bill/coin $a = (1,2,5)$ and hard to decompose if it requires more than one item $b = (3,4,6,7,8,9)$. The exercises will be generated by choosing prices with these properties and picking an object that is priced realistically. A dimension related to the difficulty is the presence of **Carried Numbers** in the operation, when there are two objects. It is managed by a different parameter because it is not related to a particular exercise type. **Price Presentation** varies due to the different practices in stores and countries, which do not always follow the standardized rule. Finally, different **Money Shapes** are used: Real Euro or poker tokens, which can reduce the visual ambiguity.

Graphical interfaces in ITS can have unwanted side effects. For this reason, the interface was entirely designed with the help of both a specialist of didactic of mathematics and primary school teacher, with several specific design choices motivated by instructional design principles, and motivational and attentional requirements. For example, the interface, shown in Figure 11, is such that: a) display is as clear and simple as possible; b) there is no chronometer, so that students are not put under time pressure; c) coins and banknotes have realistic visual appearance, and their relative sizes are respected; d) costumer and

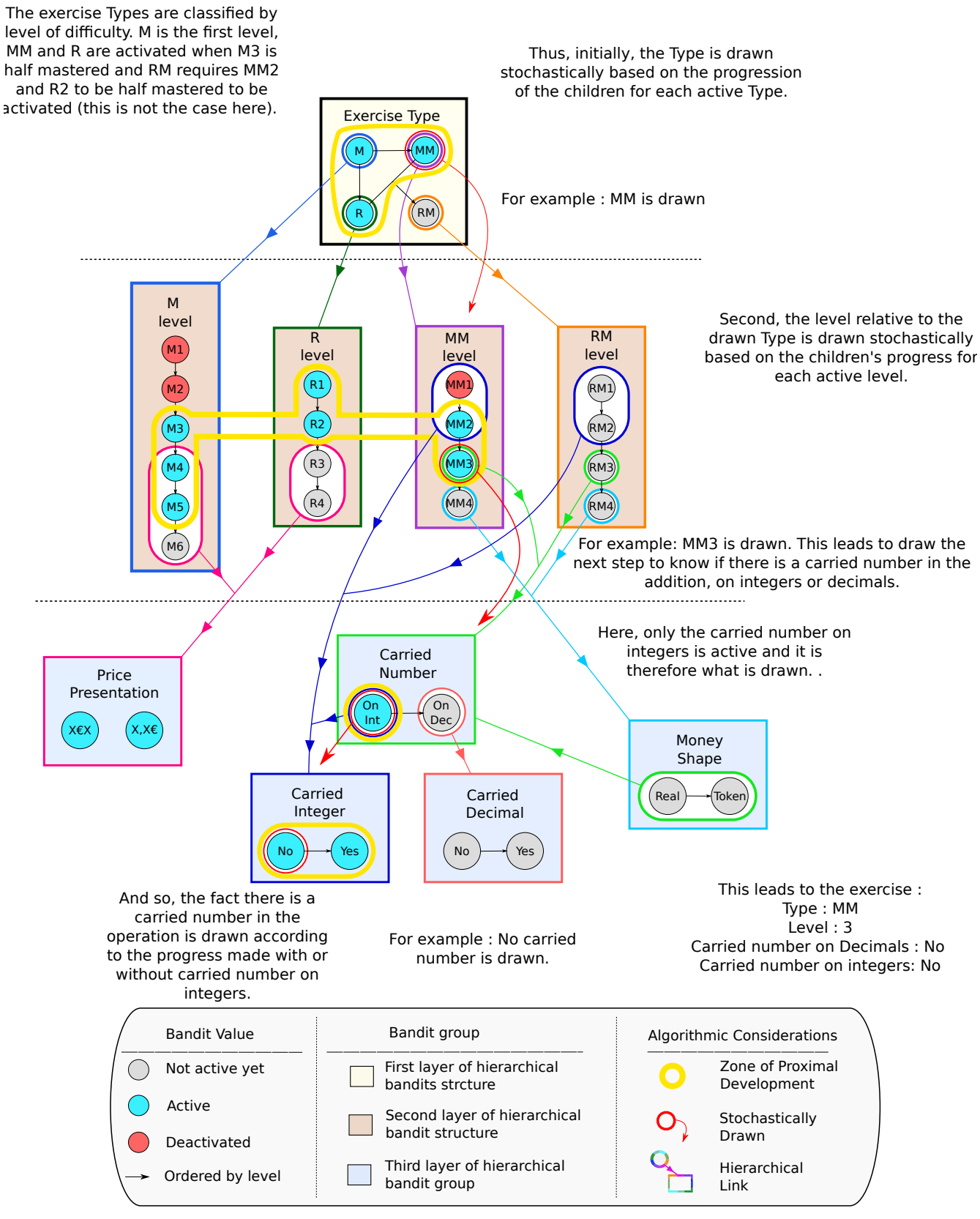


Figure 12. A representation of the activity graph used for a pedagogical scenario to teach child about mathematics by making them manipulate money. An example of activity is sampled at a particular state of the algorithm resulting from activities already made by a hypothetical student.

merchant are represented to indicate clearly the role of the student; e) text quantity is kept to minimum.

Regarding the interactions with Kidlearn game, the activity starts either with or without a choice for the student between two objects or two groups of objects as in figure 10, then one or two objects with their respective price are shown.

To complete the exercise, the student has to drag and drop the money that she/he wants to use from the wallet location to the repository location. It is possible to request extra cues, by clicking on the smiley with a hat. They have to click on the "OK" button to submit the answer leading to a feedback. If the answer is correct, the feedback is "Congratulation you can move on to the next exercise". The experience must provide the most pedagogical gains and so, the student has 3 opportunities to solve the exercise and extra cues are provided each time the student makes a try. If after 3 trials the answer is still wrong, then a feedback with the correct solution is given and the system moves on to the next exercise.

The Kidlearn game runs on tablet. The equipment was composed of a set of 30 tablets, two computers and two wifi routers. This equipment was carried in every participating classroom to be independent from school equipment constraints and limit equipment bias.

5.6 Experimental protocol

The experiment is a Randomized and Controlled Trial (RCT) approved by Inria COERLE, the lab ethical committe, which verified that all experiments are performed in accordance with French regulations as well as ethic considerations.

5.6.1 Experiment design

According to the RCT methods, the experiment was designed to compare four experimental conditions corresponding to two manipulations, i.e. the algorithm conditions, and the object choice and non-choice conditions. A predefined sequence (called "Predef") is used as a algorithm baseline with a linear learning path. It is implemented as a series of activities in which the student must have 75% success over 4 activities of the same type to pass to the next activity type. This Predef sequence was designed by a professional in didactics of mathematics. The other condition is the ZPDES algorithm condition with personalized paths according to learning progress of student (using a parameterization of learning activities previously designed by the expert). Within each algorithm condition, we introduced the possibility to self-choose the object of the money exchange for manipulating intrinsic motivation thanks to self-decision making. Hence, four conditions were manipulated as follows: 1) a "Predef" condition with linear path and without self-choice ; 2) a "PCO" condition with linear path and with self-choice of objects; 3) a "ZPDES" condition with Learning-progress based personalizing, but without self-choice of objects ; and finally 4) a "ZCO" condition with Learning-progress based personalizing with self-choice of objects. Thanks to these 4 condition, we were able to assess the effect of Learning-progress based personalizing on learning and motivation performance as well as its possible synergistic effect with playful feature related to the self-choice of object for exercise.

5.6.2 Participants

Teachers from 11 primary schools with 2nd grade classes signed up to participate in the Kidlearn program in Nouvelle-Aquitaine, the South-West region of France. Additionally, we collected the consent from each participant and their parents. To assign each student to one of the 4 experimental conditions, randomization was conducted at the classroom level in order to avoid contamination effects. Although 414 students took the pre-intervention assessment, data from 147 students were excluded from the analysis as they did not complete all the sessions of the experimental protocol (Kidlearn session and/or post-test assessment). In addition, to ensure balance across conditions ex ante, we performed a pseudo-randomized selecting procedure at student level via a computer algorithm. In particular, children were first partitioned into strata according to the following variables: the gender (girl or boy), the child age (7 or 8 years old) and pre-assessment calculation score. Then, within each stratum, children were randomly assigned to a condition.

The final sample includes 265 children in 24 classes in 11 schools with 62 children for Predef, 59 children for PCO, 76 children for ZPDES and 68 children for ZCO condition. The background characteristics of the students in terms of demographic and school-related dimensions are summarized in table 4 in appendix.

5.6.3 Measurement toolkit

The measurement toolkit included two main parts of measurement. The former referred to a profile assessment and the latter to a Kidlearn intervention assessment. A timeline was designed to articulate these assessments around the Kidlearn game inside each session (see table 3).

Profile metrics To have an assessment of the background of each participant, several measures have been collected. First, the General Profile (GP) questionnaire refers to questions related to student information such as gender, experience with technologies, his perception and habits to use money (simple manipulation or money calculation). Another set of questions concerns the habits of choosing in life-related decision making such as clothes or food choices (self-choice score): it is used to

probe the student's self-determination trait[1]. Second, a school-related psychological perception assessment has been carried from the use of two questionnaires,i.e., the Quality of School-Life Scale (QSLS)[81] and the Learner Empowerment Scale (LES)[82]. The QSLS evaluates the quality of school life experienced by the student (satisfaction at school, the student's interest in academic learning, and the nature of the student-teacher interactions / students' attitude towards the teacher), and it's a high predictor of disengagement behaviours in school. The LES measures the learner's empowerment with questions such as "This course will help me achieve my future goals" and "I have the qualifications to succeed in this class". The two questionnaires were combined and reworked to produce the School Profile questionnaire (SP), which contains 10 items on a 5-point Likert scale ; 5 items are related to school while the others to relationships.

The overall profile metrics aimed to establish an initial profile for each student in order to have equivalent control and experimental groups in respect of personal factors that may affect the results of our experiment such as demographic factors, technology experience, everyday self-determined behaviors, money manipulation and calculation and school experience and perceptions (see table 4 in appendix).

Kidlearn Intervention Assessment To assess Kidlearn Intervention according to the 4 experimental conditions, the assessment was entailed three parts. The first one corresponded to learning activity data from direct interactions with the KidLearn game computed as an "Activity score". The second part is dedicated to the learning effectiveness assessment of the Kidlearn intervention with pre-and post-test regarding calculation performance before and after the Kidlearn intervention. Finally, the third part included assessments regarding the learning experience elicited across the sessions of Kidlearn intervention (emotional and motivation scales for probing the learning experience according to the four manipulated conditions)

Activity Score from KidLearn Game The exploitation of the data from the interaction of the students with the Kidlearn activity is done using two kinds of indicators tracing the learning path of each student.

The first one (Fig. 3) is used to compare quantitatively the activities made by the students. Here, two scores are built, and used in result section 2.1. The former represents the activities reached by a student in the activity space, and the latter one represents the success rate over these reached activities. For a student, these scores are defined as follows:

$$score^{reached}(t) = \sum_{i=1}^{4} max(\{l^{i,j}(t) \mid j \in L^i\}) f^i \tag{4}$$

$$score^{success}(t) = \sum_{i=1}^{4} max(\{\delta_{i,j}(t) l^{i,j}(t) \mid j \in L^i\}) f^i \tag{5}$$

where i is the index corresponding to each type of activity, f^i is the factor related to the activity type i as described in table 2. If the level j for type i has been reached at time t, then $l^{i,j}(t) = j$, or else $l^{i,j}(t) = 0$. $\delta_{i,j}(t)$ is the student's success rate over the 4 last steps activity type i, level j. For example, at time t, if a student has reached M4, MM2, R1 and did not reach RM, his $score^{reached}$ is equal to : $4 \times 1 + 2 \times 2 + 1 \times 3 + 0 \times 4 = 11$.

	M	MM	R	RM
Index i	1	2	3	4
Factor f^i	1	2	3	4
Levels L^i	1-6	0-4	0-4	0-4

Table 2. Table of factor, index and the number of levels for each type of activity to compute scores in equations 4 and 5. The level 0 represents the fact a student has not made any exercise of this type yet. Students start with exercises of type M, so there is no level 0

The second indicator (Fig.4) dynamically traced the curriculum of each student in terms of activities performed during Kidlearn sessions. Precisely, for a given time step t and condition, a matrix slot represents the state of an activity (ordinate) for a particular student (abscissa). A slot is coded in grey if a student has never explored the corresponding activity and it is coded in purple if the student is doing this activity at time t. When a student has explored an activity, the slot is coded of tint of green depending on the student's success rate (light green: low, dark green: high).

[1]When a behaviour is self-determined, the regulatory process is choice, but when it is controlled, the regulatory process is compliance (or in some cases defiance)[80].

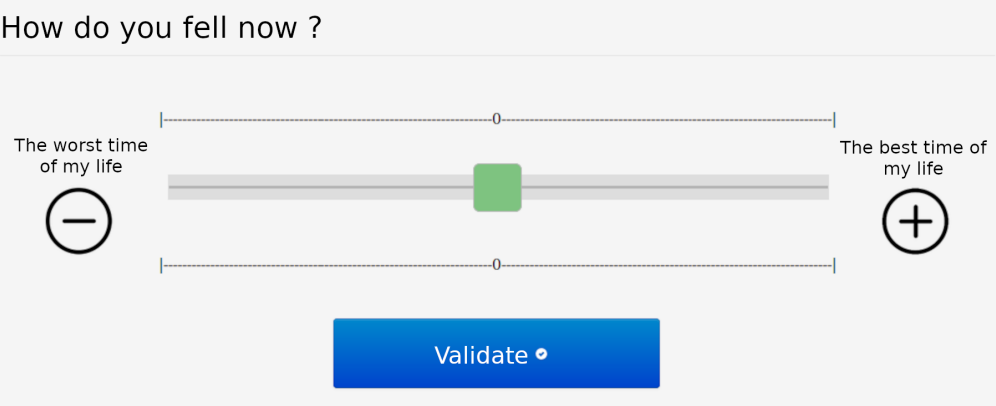


Figure 13. Emotional scale. The student is asked a question “How do you feel ?” and moves towards + (best time) or - (worst time) depending on how he/she feels right now.

KidLearn Learning Score (pre-post effect) A math-test, used as a pre-post metric to measure the students money-related computational learning, has been made by a pedagogical expert. The pre- and post-test are composed of 20 items scoring from 0 to 1 (max score is 20). The three first questions are general mathematical questions about composition of numbers. The 17 other questions are related to the manipulation money, composition of numbers, addition and subtraction, and are directly related to the activities of the Kidlearn scenario. The pre-test happens at the beginning of the first session, while the post-test happens at the end of the last session. The pre- and post-test are presented on the tablet on a dedicated interface (different from the ITS one). Each item of the test evaluates the student over knowledge and skills related to money manipulation, number composition, addition or subtraction (similar to the skills and knowledge trained in the ITS). Both tests include the same items organised in the same order but the items’ wording have randomly selected values for each item and each student (with verification that no items in the post-test have the same values in their wording as the ones in the pre-test for one student).

This way, each question is generated with a unique structure but with random values set to have the same mathematical difficulty with the goal to reduce learning effect (test-retest effect) on learning score due to the test repetition (pre-post test). This also allowed to reduce the possibilities of student cheating.

Kidlearn-related Learning experience Two questionnaires rates learning experience from Kidlearn game, using an emotional scale and a motivation questionnaire.

The first one assessed the emotional experience or well-being related to Kidlearn practice. It consisted of self-measurements of emotional valence elicited during each Kidlearn session. This self-measurement is a simplified version of the Self-Assessment Manikin[83]. It is presented in figure 13. The student must position a cursor between "the best time of my life" (intensive and positive emotional state) and "the worst time of my life" (intensive and negative emotional state) depending on how he/she feels right now (it is scored from -50 to 50). They have to answer to this scale at the beginning, middle and end of each one of the four sessions. A summative score is computed from all inputs (used in result section 2.3), max score is 600 across the four sessions. This provides us an approximate measure of the evolution of the well-being of the children during each session.

The second measure is a motivation questionnaire from Vallerand’s questionnaire[50,51] providing a measure of the amount of student motivation. It is based on Self-Determination Theory[52] and is commonly used to ascertain the elicitation of intrinsic and extrinsic motivation[53]. Once adapted to a public of children users playing a serious game, children had to answer a 21 items-questionnaire on their experience playing during the last session (max score of global motivation is 21).

All this measurement toolkit was integrated into the interface of the web application used in the experiment. Hence, no paper-pencil assessment was used, and then the data collecting was fully computerized to standardize the collecting procedure and to simplify after both the data base structuring and its statistical analysis .

5.6.4 Study Procedure

The study procedure followed several steps managed by two experimenters. Precisely, four experimental sessions are organized over two weeks where the three Kidlearn game phases (30 mn/phase), the profile metrics as well as Kidlearn related assessment were performed. Each participant had to undergo the four successive sessions according to a specific organisation of its contents. The organisation of contents of each session is presented in table 3.

There are some points to consider in particular. A game phase is always preceded by a self-emotional assessment and followed by a self-emotional assessment scale. The questionnaires were distributed over the 4 sessions so as not to ask too

Session 1 (∼1h20)	Session 2 (∼40 min)	Session 3 (∼40 min)	Session 4 (∼40min)
1. Project explanation	1. Emotional scale	1. Emotional scale	1. Emotional scale
2. Emotional scale	2. Game phase (30 min)	2. SP questionnaire	2. Motivation questionnaire
3. GP questionnaire	3. Emotional scale	3. Emotional scale	3. Emotional scale
4. Math pre-test (20 min)	4. GI questionnaire	4. Game phase (30 min)	4. Math post-test (20 min)
5. Emotional scale	5. Emotional scale	5. Emotional scale 3	5. Emotional scale
6. Game phase (30 min)			
7. Emotional scale			

Table 3. Sessions Timeline. The table shows the sequence of steps for each session. There are 7 steps in the first session and 5 steps in the other sessions. The GI questionnaire is a hand made questionnaire done to evaluate our Kidlearn interface, results of this questionnaire is not include in the study due to technical problem in collecting data.

many questions at a time. In the first session, the General Profile (GP) questionnaire is done first. They have all the time they want to answer. It allows students to acclimatize to the tablet, the web site interface and to be confident in the tablet usage. After, the Math pre-test followed by a game phase were administrated. The session 2 and 3 are similarly structured except the exclusion of contents related to project explanation and GP profile. Finally, the session 4 is dedicated to post intervention assessment including the motivation questionnaire and the Math post-test.

For providing an optimal learning environment for the experiment, each class group was divided into two sub-groups for the whole of sessions.Each sub-group was doing the same session in parallel in different rooms supervised by a researcher. At each step of the session, finishing a step leads the student to the waiting page. While waiting for the others to finish the step, students can draw on their draft or if waiting time may be long (as for tests, some finished what they can do in 10 minutes), they can read a book present in the classroom or discuss with other that also finished without disturbing the classroom. This process allowed to considerably reduce distraction and provided a better and quieter working environment for the students.

6 Acknowledgements

We thank the teachers of the Bordeaux School District for providing access to their classrooms, an essential component for our empirical study, as well as the Rectorate of Bordeaux-Nouvelle Aquitaine, a regional administrative division of the French Ministry of Education, and its Digital Education mission, with whom Inria established a partnership convention, enabling our access to educational institutions. Special acknowledgment is given to Josias Levi Alvarez and Medhi Alaimi, two internship students, who were pivotal in conducting the in-class experiments, significantly contributing to the success of our study. This project benefited from funding from ANR AI Chair DeepCuriosity ANR-19-CHIA-0004.

7 Appendix

7.1 Population characteristics for each condition

Table 4 shows the student's frequencies for gender and calculation liking as well as the mean and standard deviation for each group (Predef, PCO, ZPDES and ZCO condition) regarding the studied profile metrics (Self-choice, Technology experience, Money manipulation, Money calculation and School-related perception). Group comparisons did not reach the significance ($p > .05$).

	Predef		PCO		ZPDES		ZCO	
Gender	F	M	F	M	F	M	F	M
	31	31	29	30	38	38	46	22
Like calculation	Yes	No	Yes	No	Yes	No	Yes	No
	53	9	54	5	68	8	53	15
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Self-choice score	9.39	3.19	8.73	2.97	8.82	2.75	9.10	2.53
Technology Experience	3.05	1.84	3.05	1.83	3.51	1.69	3.19	1.73
Money Manipulation	2.45	1.00	2.37	0.89	2.38	0.89	2.50	0.82
Money Calculation	4.16	1.23	4.16	1.08	4.17	1.11	4.06	1.26
School Profile	27.32	6.15	27.32	6.79	28.53	6.11	28.62	5.53

Table 4. Population characteristics for each condition

7.2 Detailed predefined sequence

Table 5 shows the 27 successive activities for the students following the parameters defined in section 5.5.

	G1.1	G1.2	G1.3	G2.1	G2.2	G2.3	G2.4	G3.1	G3.2	G4.1	G4.2	G4.3	G4.4
Ex Type	M	M	M	MM	MM	MM	MM	R	R	RM	RM	RM	RM
Difficulty	1	2	3	1	1	2	2	1	2	1	1	2	2
Cents Not	-	-	-	-	-	-	-	-	-	-	-	-	-
Remainder	-	-	-	-	-	-	-	-	-	-	Int	-	Int
Money Type	Real	Real	Real	Real	Real	Real	Real	Real	Real	Real	Real	Real	Real

	G5.1	G5.2	G5.3	G5.4	G6.5	G6.6	G6.7	G6.8	G7.1	G7.2	G7.3	G8.5	G8.6	G8.7	G8.8
Ex Type	M	M	M	M	MM	MM	MM	MM	R	R	R	RM	RM	RM	RM
Difficulty	4	5	5	6	3	3	4	4	3	3	4	3	3	4	4
Cents Not	x€x	x€x	x,x€	x,x€	-	-	-	-	x€x	x€x	x,x€	-	-	-	-
Remainder	-	-	-	-	-	Int	-	Dec	Int	-	Int	-	-	Int	Dec
Money Type	Real	Real	Real	Real	Real	Real	Real	Token	Real	Real	Real	Real	Real	Real	Token

Table 5. Detailed predefined sequence

References

1. Bartolomé, A., Castañeda, L. & Adell, J. Personalisation in educational technology: The absence of underlying pedagogies. *International journal of educational technology in higher education* **15**, 1–17 (2018).
2. Deunk, M. I., Doolaard, S., Smalle-Jacobse, A. & Bosker, R. J. *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices* (GION onderwijs/onderzoek, Rijksuniversiteit Groningen, 2015).
3. Iterbeke, K., De Witte, K. & Schelfhout, W. The effects of computer-assisted adaptive instruction and elaborated feedback on learning outcomes. a randomized control trial. *Computers in Human Behavior* **120**, 106666 (2021).
4. Collins, A. & Halverson, R. *Rethinking education in the age of technology: The digital revolution and schooling in America* (Teachers College Press, 2018).
5. Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. Cognitive tutors: Lessons learned. *The journal of the learning sciences* **4**, 167–207 (1995).
6. Koedinger, K. R., Anderson, J. R., Hadley, W. H., Mark, M. A. *et al.* Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* **8**, 30–43 (1997).
7. Nkambou, R., Mizoguchi, R. & Bourdeau, J. *Advances in intelligent tutoring systems*, vol. 308 (Springer, 2010).
8. Vandewaetere, M., Desmet, P. & Clarebout, G. The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior* **27**, 118–130 (2011).
9. Sun, S., Joy, M. & Griffiths, N. The use of learning objects and learning styles in a multi-agent education system. *Journal of Interactive Learning Research* **18**, 381–398 (2007).
10. Bunderson, C. V. & Martinez, M. Building interactive world wide web (web) learning environments to match and support individual learning differences. *Journal of Interactive Learning Research* **11**, 163–195 (2000).
11. Koedinger, K. R. & Anderson, J. R. Effective use of intelligent software in high school math. In *Proceedings of AI-ED 93, World conference of Artificial Intelligence in Education*, 241–248 (1993).
12. Ray, R. D. & Belden, N. Teaching college level content and reading comprehension skills simultaneously via an artificially intelligent adaptive computerized instructional system. *The Psychological Record* **57**, 201–218 (2007).
13. Milne, S., Cook, J., Shiu, E. & McFadyen, A. Adapting to learner attributes: Experiments using an adaptive tutoring system. *Educational psychology* **17**, 141–155 (1997).
14. Conati, C., Gertner, A. & Vanlehn, K. Using bayesian networks to manage uncertainty in student modeling. *User modeling and user-adapted interaction* **12**, 371–417 (2002).
15. Vygotsky, L. S. *Mind in society: The development of higher psychological processes* (Harvard university press, 1930-1934/1978).
16. Csikszentmihalyi, M. & Csikszentmihalyi, I. *Beyond boredom and anxiety*, vol. 721 (Jossey-Bass San Francisco, 1975).
17. Sweller, J. Cognitive load theory. In *Psychology of learning and motivation*, vol. 55, 37–76 (Elsevier, 2011).
18. Murayama, K. A reward-learning framework of autonomous knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards. *OSF Preprints* **10** (2019).
19. Kaplan, F. & Oudeyer, P.-Y. In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience* **1**, 225 (2007).
20. Oudeyer, P.-Y., Gottlieb, J. & Lopes, M. Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In *Progress in brain research*, vol. 229, 257–284 (Elsevier, 2016).
21. Gottlieb, J., Oudeyer, P.-Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences* **17**, 585–593, DOI: 10.1016/j.tics.2013.09.001 (2013).
22. Ryan, R. M. & Deci, E. L. Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology* **61**, 101860 (2020).
23. Ten, A., Kaushik, P., Oudeyer, P.-Y. & Gottlieb, J. Humans monitor learning progress in curiosity-driven exploration. *Nature communications* **12**, 5972 (2021).
24. Oudeyer, P.-Y., Kaplan, F. & Hafner, V. V. Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation* **11**, 265–286 (2007).
25. Schmidhuber, J. Curious model-building control systems. In *Neural Networks, 1991. 1991 IEEE International Joint Conference on*, 1458–1463 (IEEE, 1991).

26. Lopes, M., Lang, T., Toussaint, M. & Oudeyer, P.-Y. Exploration in model-based reinforcement learning by empirically estimating learning progress. In *Advances in Neural Information Processing Systems*, 206–214 (2012).

27. Graves, A., Bellemare, M. G., Menick, J., Munos, R. & Kavukcuoglu, K. Automated curriculum learning for neural networks. In *international conference on machine learning*, 1311–1320 (PMLR, 2017).

28. Colas, C., Fournier, P., Chetouani, M., Sigaud, O. & Oudeyer, P.-Y. Curious: intrinsically motivated modular multi-goal reinforcement learning. In *International conference on machine learning*, 1331–1340 (PMLR, 2019).

29. Kim, K., Sano, M., De Freitas, J., Haber, N. & Yamins, D. Active world model learning with progress curiosity. In *International conference on machine learning*, 5306–5315 (PMLR, 2020).

30. Portelas, R., Colas, C., Weng, L., Hofmann, K. & Oudeyer, P.-Y. Automatic curriculum learning for deep rl: A short survey. In *IJCAI* (2020).

31. Clément, B., Roy, D., Oudeyer, P.-Y. & Lopes, M. Online Optimization of Teaching Sequences with Multi-Armed Bandits. In *7th International Conference on Educational Data Mining* (London, United Kingdom, 2014).

32. Auer, P., Cesa-Bianchi, N., Freund, Y. & Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing* **32**, 48–77 (2003).

33. Bubeck, S. & Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Stochastic Systems* **1** (2012).

34. Clément, B., Roy, D., Oudeyer, P.-Y. & Lopes, M. Multi-Armed Bandits for Intelligent Tutoring Systems. *Journal of Educational Data Mining (JEDM)* **7**, 20–48 (2015).

35. Clément, B., Oudeyer, P.-Y. & Lopes, M. A Comparison of Automatic Teaching Strategies for Heterogeneous Student Populations. In *EDM 16 - 9th International Conference on Educational Data Mining*, Proceedings of the 9th International Conference on Educational Data Mining (Raleigh, United States, 2016).

36. Leotti, L. A. & Delgado, M. R. The inherent reward of choice. *Psychological science* **22**, 1310–1318 (2011).

37. Murayama, K. *et al.* How self-determined choice facilitates performance: A key role of the ventromedial prefrontal cortex. *Cerebral Cortex* **25**, 1241–1251 (2013).

38. Cordova, D. I. & Lepper, M. R. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology* **88**, 715 (1996).

39. Byrnes, J. P. *The nature and development of decision-making: A self-regulation model* (Psychology Press, 2013).

40. Miller, D. C. & Byrnes, J. P. To achieve or not to achieve: A self-regulation perspective on adolescents' academic decision making. *Journal of Educational Psychology* **93**, 677 (2001).

41. Wigfield, A. & Eccles, J. S. The development of achievement task values: A theoretical analysis. *Developmental review* **12**, 265–310 (1992).

42. Brophy, J. Toward a model of the value aspects of motivation in education: Developing appreciation for.. *Educational psychologist* **34**, 75–85 (1999).

43. Harter, S. A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental psychology* **17**, 300 (1981).

44. Aleven, V., McLaughlin, E. A., Glenn, R. A. & Koedinger, K. R. Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction* 522–560 (2016).

45. Faber, J. M., Luyten, H. & Visscher, A. J. The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & education* **106**, 83–96 (2017).

46. Ma, W., Adesope, O. O., Nesbit, J. C. & Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology* **106**, 901 (2014).

47. Cheung, A. C. & Slavin, R. E. The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational Research Review* **9**, 88–113, DOI: https://doi.org/10.1016/j.edurev.2013.01.001 (2013).

48. Hew, K. F. & Cheung, W. S. Use of web 2.0 technologies in k-12 and higher education: The search for evidence-based practice. *Educational research review* **9**, 47–64 (2013).

49. Gerard, L., Matuk, C., McElhaney, K. & Linn, M. C. Automated, adaptive guidance for k-12 education. *Educational Research Review* **15**, 41–58 (2015).

50. Vallerand, R. J. *et al.* The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and psychological measurement* **52**, 1003–1017 (1992).

51. Vallerand, R. J., Blais, M. R., Brière, N. M. & Pelletier, L. G. Construction et validation de l'échelle de motivation en éducation (eme). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* **21**, 323 (1989).

52. Deci, E. L. & Ryan, R. M. The general causality orientations scale: Self-determination in personality. *Journal of research in personality* **19**, 109–134 (1985).

53. Desrochers, A., Comeau, G., Jardaneh, N. & Green-Demers, I. L'élaboration d'une échelle pour mesurer la motivation chez les jeunes élèves en piano. *Recherche en éducation musicale* **24**, 13–33 (2006).

54. Proulx, J.-N., Romero, M. & Arnab, S. Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning. *Simulation & Gaming* **48**, 81–97 (2017).

55. Tyack, A. & Mekler, E. D. Self-determination theory in hci games research: current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–22 (2020).

56. Bitterly, T. B., Mislavsky, R., Dai, H. & Milkman, K. L. Dueling with desire: a synthesis of past research on want/should conflict. *Should Conflict (February 28, 2014)* (2014).

57. Grund, A., Grunschel, C., Bruhn, D. & Fries, S. Torn between want and should: An experience-sampling study on motivational conflict, well-being, self-control, and mindfulness. *Motivation and Emotion* **39**, 506–520 (2015).

58. Bernecker, K. & Ninaus, M. No pain, no gain? investigating motivational mechanisms of game elements in cognitive tasks. *Computers in Human Behavior* **114**, 106542 (2021).

59. Alamri, H., Lowell, V., Watson, W. & Watson, S. L. Using personalized learning as an instructional approach to motivate learners in online higher education: Learner self-determination and intrinsic motivation. *Journal of Research on Technology in Education* **52**, 322–352 (2020).

60. Roediger III, H. L. & Karpicke, J. D. The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science* **1**, 181–210 (2006).

61. Mazon, C., Clément, B., Roy, D., Oudeyer, P.-Y. & Sauzéon, H. Pilot study of an intervention based on an intelligent tutoring system (its) for instructing mathematical skills of students with asd and/or id. *Education and Information Technologies* (2022).

62. Delmas, A., Clement, B., Oudeyer, P.-Y. & Sauzéon, H. Fostering health education with a serious game in children with asthma: pilot studies for assessing learning efficacy and automatized learning personalization. In *Frontiers in Education*, 99 (Frontiers, 2018).

63. Clancey, W. J. Acquiring, representing, and evaluating a competence model of diagnostic strategy.(rep. no. stan-cs-1067.) stanford, ca: Department of computer science (1985).

64. Levesque, H. J. Knowledge representation and reasoning. *Annual review of computer science* **1**, 255–287 (1986).

65. Russell, S. J., Norvig, P. & Davis, E. Upper saddle river. *Artificial intelligence: a modern approach. 3rd ed. Prentice Hall: NJ* (2009).

66. Aleven, V. & Koedinger, K. R. Knowledge component (kc) approaches to learner modeling. *Design recommendations for intelligent tutoring systems* **1**, 165–182 (2013).

67. Sottilare, R. A. *et al. Design recommendations for intelligent tutoring systems: Volume 4-domain modeling*, vol. 4 (US Army Research Laboratory, 2016).

68. Nájera, P., Sorrel, M. A., de la Torre, J. & Abad, F. J. Balancing fit and parsimony to improve q-matrix validation. *British Journal of Mathematical and Statistical Psychology* **74**, 110–130 (2021).

69. Clément, B. *Adaptive Personalization of Pedagogical Sequences using Machine Learning*. Theses, Université de Bordeaux (2018).

70. Oudeyer, P.-Y. & Kaplan, F. What is intrinsic motivation? a typology of computational approaches. *Frontiers in neurorobotics* **1** (2007).

71. Cesa-Bianchi, N. *et al.* How to use expert advice. *Journal of the ACM (JACM)* **44**, 427–485 (1997).

72. Lopes, M. & Oudeyer, P.-Y. The strategic student approach for life-long exploration and learning. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, 1–8 (IEEE, 2012).

73. Basawapatna, A. R., Repenning, A., Koh, K. H. & Nickerson, H. The zones of proximal flow: guiding students through a space of computational thinking skills and challenges. In *Proceedings of the ninth annual international ACM conference on International computing education research*, 67–74 (ACM, 2013).

74. Metcalfe, J., Schwartz, B. L. & Eich, T. S. Epistemic curiosity and the region of proximal learning. *Current opinion in behavioral sciences* **35**, 40–47 (2020).

75. Fournier-Viger, P., Nkambou, R., Nguifo, E. M. & Mayers, A. Its in ill-defined domains: toward hybrid approaches. In *International Conference on Intelligent Tutoring Systems*, 318–320 (Springer, 2010).

76. Bloom, B. S. Learning for mastery. instruction and curriculum. regional education laboratory for the carolinas and virginia, topical papers and reprints, number 1. *Evaluation comment* **1**, n2 (1968).

77. Roy, D. *Usage d'un robot pour la remédiation en mathématiques*. Master's thesis, Université de Bordeaux (2012).

78. Lopes, M. & Oudeyer, P.-Y. The strategic student approach for life-long exploration and learning. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, 1–8 (IEEE, 2012).

79. Carstensen, L. L., Fung, H. H. & Charles, S. T. Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and emotion* **27**, 103–123 (2003).

80. Deci, E. L., Vallerand, R. J., Pelletier, L. G. & Ryan, R. M. Motivation and education: The self-determination perspective. *Educational psychologist* **26**, 325–346 (1991).

81. Lazar, S. M. *The Quality of School Life Scale as a predictive indicator of student disengagement from school* (1999).

82. Weber, K., Martin, M. M. & Cayanus, J. L. Student interest: A two-study re-examination of the concept. *Communication Quarterly* **53**, 71–86 (2005).

83. Bradley, M. M. & Lang, P. J. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* **25**, 49–59 (1994).