



**HAL**  
open science

# Using Prompts and Remediation to Improve Primary School Students Self-Evaluation and Self-Efficacy in a Literacy Web Application

Thomas Sergent, François Bouchet, Morgane Daniel, Thibault Carron

► **To cite this version:**

Thomas Sergent, François Bouchet, Morgane Daniel, Thibault Carron. Using Prompts and Remediation to Improve Primary School Students Self-Evaluation and Self-Efficacy in a Literacy Web Application. 16th European Conference on Technology Enhanced Learning, Sep 2021, Bolzano, Italy. pp.221-234, 10.1007/978-3-030-86436-1\_17. hal-03353767

**HAL Id: hal-03353767**

**<https://hal.science/hal-03353767>**

Submitted on 24 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Using Prompts and Remediation to Improve Primary School Students Self-Evaluation and Self-Efficacy in a Literacy Web Application

Thomas Sergent<sup>1,2</sup>[0000-0003-0908-4258], François Bouchet<sup>1</sup>[0000-0001-9436-1250],  
Morgane Daniel<sup>2</sup>, and Thibault Carron<sup>1</sup>[0000-0001-6982-7055]

<sup>1</sup> Sorbonne Université, CNRS, LIP6, F-75005 Paris, France  
{firstname.lastname}@lip6.fr

<sup>2</sup> Lalilo, Paris, France {firstname}@lalilo.com

**Abstract.** Self-regulation skills are critical for students of all ages in order to maximize their learning. A key aspect of self-regulation is being aware of one's performance and deficits in self-evaluation. Additionally, a clear consensus has not been reached regarding the age one can start learning these self-regulation processes. In order to investigate the possibility to raise awareness to some self-regulation deficits in 5 to 8 years old children, we have introduced two prompts triggered randomly after 1 out of 15 exercises into a literacy web-application for primary school students, to evaluate perceived difficulty [Too easy, Good, Too difficult] and desired difficulty [easier, same level, harder]. Comparing students' actual performance with their responses to self-regulatory prompts can provide information about their ability to self-regulate their learning, in particular in terms of self-evaluation and self-efficacy. We collected 2,600,142 responses from 467,116 students for our experiments. The goal of this paper is to assess the impact of two different remediation strategies to reduce the two types of deficits initially measured in students.

In a first study, we measured the impact of a gauge (resp. an audio recording) showing (resp. telling) the number of correct and incorrect answers to help students evaluate their actual performance during answers to the self-regulation prompts. In a second study, we measured the impact of giving self-evaluation and self-efficacy remediation to students who showed a deficit in self-regulated learning abilities from their answers to the self-regulation prompts.

The results show (a) a significant reduction of self-evaluation deficits when answers were supported by a visual gauge, (b) no significant impact on self-evaluation deficits when answers were supported by an audio recording, (c) a significant reduction of future self-evaluation deficits when giving students audio feedback advising them not to repeat a detected deficit.

This underlines the possibility of scaffolding self-regulated learning skills in a web based application from a young age while learning another skill.

**Keywords:** Self-Regulated Learning · Primary School · Web Based Application · Scaffolding · Remediation · Self-Evaluation · Self-Efficacy

## 1 Introduction

Children’s self-regulated learning (SRL) skills are a key component of their academic performance, as self-regulated students generally know better “how to learn”, which can have a positive impact in all disciplines [15]. The earlier children begin to develop these skills, the greater the impact on their overall schooling, and self-regulation training programs for elementary school students have already been developed for this purpose [4]. Nevertheless, it can be difficult for teachers to provide individualized help to each student, both in terms of the task at hand (e.g., learning to read) and in terms of their self-regulation skills. The SRL scaffolding using computer tools has also been studied: a meta-analysis of SRL support implemented up to 2016 shows their significant positive effect on progression [14]. However, these tools have several limits: firstly, they only targeted older students (beyond 5<sup>th</sup> grade), secondly they focused on the performance phase (one of the three phases of the SRL cycle described, along with the anticipation and self-reflection phases [15]), and thirdly they measured whether the student’s progression was improved by the SRL support, rather than whether the student was improving their SRL skills. Thus, SRL is mostly seen as supporting learning, not as a skill to be evaluated and trained in itself.

Previous work has shown that among young students, self-evaluation and self-efficacy deficits are two prevalent issues [12]. In this paper, our goal is to investigate how to train self-evaluation and self-efficacy through scaffolding (helping the student to assess their level) and feedback (suggesting what to do in the future), and to measure the impact of these two strategies from a data analysis point of view. More particularly, we investigate the following research questions:

(RQ1) Can scaffolding help students in correcting their self-evaluation deficits?

(RQ2) Can a remediation feedback help students in answering to future self-evaluation and self-regulation prompts?

The remainder of this paper is organized as follows: in section 2 we will present briefly some related work on self-regulation support in particular for younger children. In section 3 we will introduce the experimental context, the designed prompts and the data collected. We will then describe the two experiments led to answer to RQ1 and RQ2 respectively in sections 4 and 5, before concluding with a discussion.

## 2 Related work

SRL is a three-phase cycle that repeats itself with each new task the learner is confronted with [15]. First the learner prepares for the task (anticipation phase), then they perform the task and can monitor their progress (performance phase). Finally, they assess the effectiveness of their learning to draw conclusions for future learning (self-reflection phase).

As previously mentioned, some SRL training programs have been shown to have a significant positive effect in primary school children [4] but outside of a computer-based context. Some works have already tried to assess the effect of self-regulatory prompts to show their positive effects on self-efficacy [10]. For instance, Müller [7] showed that prompting university students had an immediate impact which did not transfer over time. Hoffman [5] also showed a positive impact

of prompting for self-efficacy but before accomplishing the task and measured the impact on performance more than in self-efficacy itself. More generally, a meta-review [9] has shown that self-assessment has a positive impact on learners self-efficacy.

It is worth noting that although young children’s abilities to use SRL strategies may be more limited than in teenagers, they seem to have comparable monitoring skills [11]. Indeed, recent work on a dashboard supporting SRL in a mathematics software program for 9-10 years old (only slightly older than our targeted students) showed a significant improvement in SRL skills for students in the dashboard group compared to those without the dashboard [6]. Finally, young students not detected as having a self-evaluation deficit seemed to rely mainly on their success rate when asked to self-evaluate [13].

### 3 Experimental context and data collection

#### 3.1 A literacy software

Lalilo is one of the many web applications used by teachers in the classroom to help them implement differentiated instruction strategies. At the beginning of 2021, it is used by 40,000 English and French speaking kindergarten and primary classes every week to strengthen literacy through series of exercises adapted to the students’ level. It also provides the teacher with a dashboard to evaluate the students’ activities and progress. It is therefore a relevant testing ground for attempting to correct students’ SRL deficits and measuring the impact of different strategies. A typical session lasts 20 minutes (on average) with the student performing around 15 short exercises with 3 to 7 questions each, chosen by an adaptive learning algorithm, as pictured by Figure 1. Student activities (e.g. logging in, time spent on a question/exercise, mistakes) are traced and we focus here only on students’ answers to an exercise, thus calling **trace** only the answers to a set of questions of the same type.

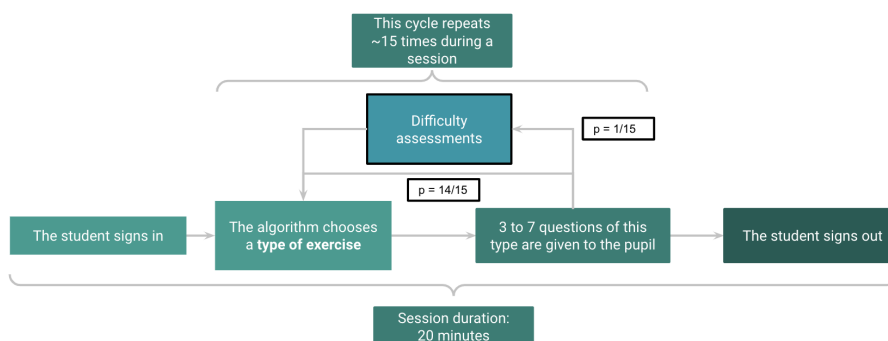


Fig. 1: Chronology of a typical student session

### 3.2 Self-evaluation and self-efficacy prompts

To assess some aspects of students’ SRL skills, we introduced two self-regulatory prompts [1] which are randomly shown successively once every fifteen exercises when a student finishes an exercise (i.e. a student answers them on average once per typical learning session). First, the **perceived difficulty** prompt asks the student *“How difficult was this exercise for you?”* with 3 possible answers: “Too hard”, “Just-right”, “Too easy”). Then, if we don’t detect any self-evaluation deficit (which are described in the next subsection), the student is asked to reply to the **desired difficulty** prompt *“I would like exercises that are...”* with 3 possible answers: “easier”, “the same level”, “harder”. The perceived difficulty prompt aims at measuring the **self-evaluation** ability of the students, i.e. their ability to correctly estimate the difficulty of the questions they just answered. The desired difficulty prompt aims at measuring their **self-efficacy**, i.e. how they would react to their representation of the difficulty. The visuals for these two prompts are displayed in Figure 2. Before introducing the assessments, we checked qualitatively in a classroom using Lalilo that prompts were understood by 1<sup>st</sup> grade students (details not presented here).

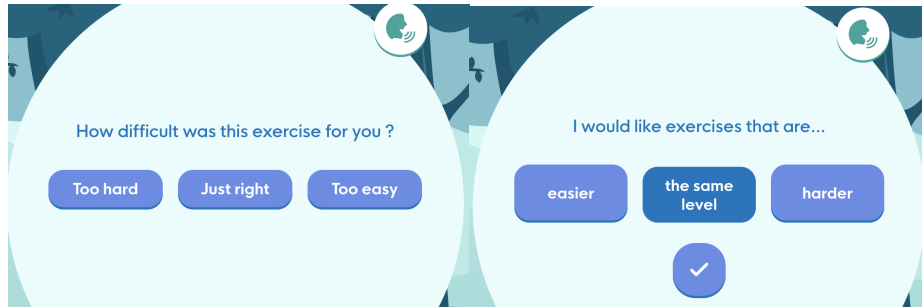


Fig. 2: Perceived (left) and desired (right) difficulty prompts

### 3.3 Data collection

We collected traces from Kindergarten, 1<sup>st</sup> grade and 2<sup>nd</sup> grade classes based in France, Canada and USA learning in French (FR) or English (EN) between January 18 and April 8, 2021 on the Lalilo platform. We kept only the traces for which students had answered to self-regulation prompts (i.e. on average 1/15<sup>th</sup> of all traces) and hereafter we call trace the answers to the exercise with the associated answers to SRL prompts.

### 3.4 Data preprocessing

**Deficit tagging** As a trace registers answers to each exercise question as well as to the two SRL prompts, one can compute the success rate of a trace defined as the number of correct answers over the total number of questions of the trace. From the success rate, we can determine a **performance tag** of a trace with one of those three values: excellent (all answers are correct), poor (34% or less of the answers are correct), and medium (for the remaining cases). We have chosen a threshold of 34% for “poor” performance so that traces that have only one correct answer out of 3 are considered poor. Indeed, the expected probability of succeeding questions is always at least at  $1/3$  which means students having a success rate of  $1/3$  or less do not perform better than chance. It is also worth noting that the “excellent” tag is quite conservative, as one could argue that a student who answered correctly to 6 out of 7 questions could be considered as having a very good performance as well.

From the performance, the perceived difficulty and the desired difficulty, we generate the so-called **trace deficit tag** displayed in Table 1. The “Desired difficulty” column is empty for the first four listed trace deficits: these deficits are self-evaluation deficits and in these cases students were not asked the desired difficulty prompt. Indeed, we considered that if the student did not have a proper representation of the difficulty of the exercise, it was not relevant to ask them the desired difficulty prompt.

Table 1: Trace deficit tag determination

Actual performance	Perceived difficulty	Desired difficulty	Deficit
excellent	too hard		underevaluation
excellent	just-right		slight underevaluation
poor	too easy		overevaluation
poor	just-right		slight overeval.
excellent	too easy	easier/same	avoiding difficulty
poor	too hard	harder/same	seeking difficulty

## 4 Impact of a gauge or an audio recording during answers to self-evaluation and self-efficacy assessments

### 4.1 Method

To answer to our first research question on how scaffolding can help with self-evaluation deficits, we focused only on the deficits involving self-evaluation only, i.e. only the four first deficits in Table 1. In order to measure the impact of visual cues on the answers to the perceived difficulty prompt (=self-evaluation prompt), students were randomly given one of two visuals for the prompts: one similar to the initial prompts (Figure 2) and one with an additional gauge displaying the number of correct and incorrect answers in the past exercise (Figure 3).

Additionally, in order to measure the impact of auditory cues, students were randomly given an audio recording stating their number of correct answers and total number of answers in their last exercise: e.g. “In the last exercise, you found three correct answers out of four questions”. This sentence is read instead of shown to not bias answers simply because younger students may not be able to read it well. For the same reason, every text displayed on the screenshots is also read aloud to the student, and they can replay the instruction using the top right-hand button. The choice of alternative modalities is therefore only because of the particular audience (young students who are not necessarily fully literate yet), and not because of an hypothesis on learning styles which have been proven to be a neuromyth [8].

Overall, when a student got a self-evaluation prompt, they were assigned randomly in one of four options: (a) no gauge and no audio recording (control condition); (b) gauge and no audio recording; (c) no gauge but audio recording; (d) gauge and audio recording. Our hypothesis was that visual or audio cues could support self-evaluation for students.

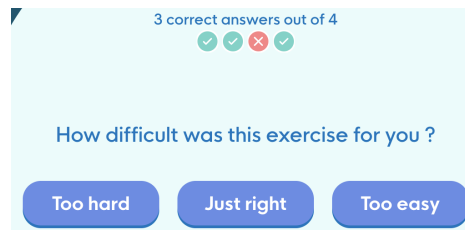


Fig. 3: The gauge shown for scaffolding, visible above both SRL prompts

## 4.2 Results

In this experiment, we only kept the first answer of students to the self-evaluation prompt so that our data would not be interfered by students answering repeatedly to the self-evaluation prompt and sometimes having visual or audio support or not. This also allows us to isolate the effect of the gauge and audio recording alone. We then selected traces with excellent performance (100% success rate as described above) and computed the answer deficit distribution using Table 1 depending on the presence of the gauge and audio recording. The results are summarized in Figure 4. We can observe that there is a significant difference in the answer deficit distribution if there is a gauge or not: students with excellent performance having a gauge as a visual support are less likely to show an underevaluation ( $d = 0.05$ ,  $p < 0.001$ , two tailed hypothesis). There is also a significant difference between the answers with an audio recording support or not: however the effect is that students with excellent performance having an audio recording are *more* likely to show an underevaluation.

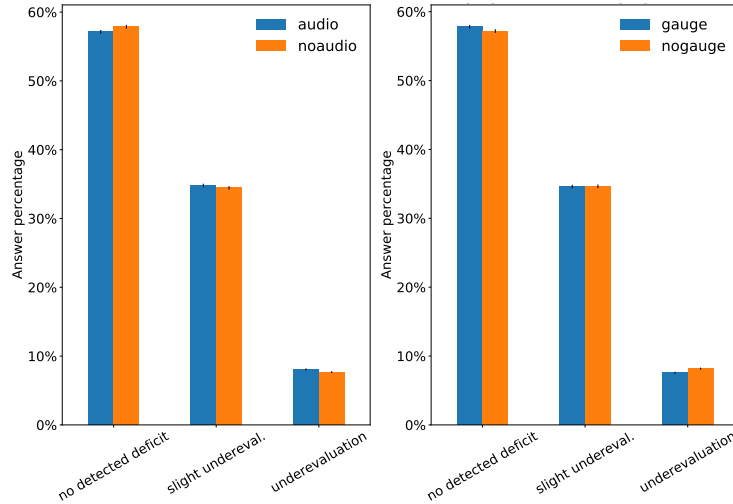


Fig. 4: Self-evaluation deficit distribution depending on gauge and audio recording presence for excellent performance answers with confidence interval at 95%.  $N(\text{“audio”}) = 136,210$ ,  $N(\text{“no audio”}) = 137,156$ ;  $N(\text{“gauge”}) = 136,716$ ,  $N(\text{“no gauge”}) = 136,650$ . No “avoiding difficulty” as it is not a self-evaluation deficit.

Similarly, we selected traces with poor performance (less than 34% success rate as described above) and computed the answer deficit distribution using Table 1 depending on the presence of the gauge and audio recording. The results are summarized in Figure 5. We can observe that there are significant differences in the answer deficit distribution both when there is a gauge ( $d = 0.06$ ,  $p < 0.001$ , two tailed hypothesis) and auditory support ( $d = 0.04$ ,  $p < 0.001$ , two tailed hypothesis), and both impact lead to less over-evaluation by students.

We can note that the baseline percentage of deficits is a lot higher for poor performance than excellent performance traces (51% against 34%). This is consistent with the fact that students who got an excellent performance are more likely to correctly self-evaluate. Moreover, the effect size - which measures the impact of the intervention - is larger for traces with poor performance than for traces with excellent performance, which indicates that students who had a poor performance seem to benefit more from visual or auditory support than those who had an excellent performance. As the impact of the gauge is positive and significant for both excellent and poor performance traces, we measured the impact of having the audio or not when there was a gauge to support a student’s self-evaluation. The results are summarized in Figure 6. They indicate that having an audio recording stating the number of correct answers over the total number of questions (e.g. “In the last exercise, you found three correct answers out of four questions”) had a significant impact on decreasing the number of student overevaluating when the students performance was poor on the last



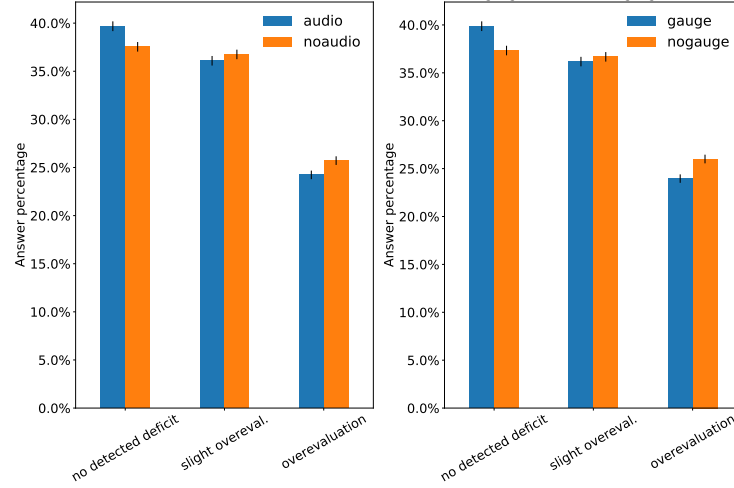


Fig. 5: Self-evaluation deficit distribution depending on gauge and audio recording presence for poor performance answers with confidence interval at 95%.  $N(\text{"audio"}) = 35,700$ ,  $N(\text{"no audio"}) = 36,394$ ;  $N(\text{"gauge"}) = 36,048$ ;  $N(\text{"no gauge"}) = 36,046$ . No “seeking difficulty” as it is not a self-evaluation deficit.

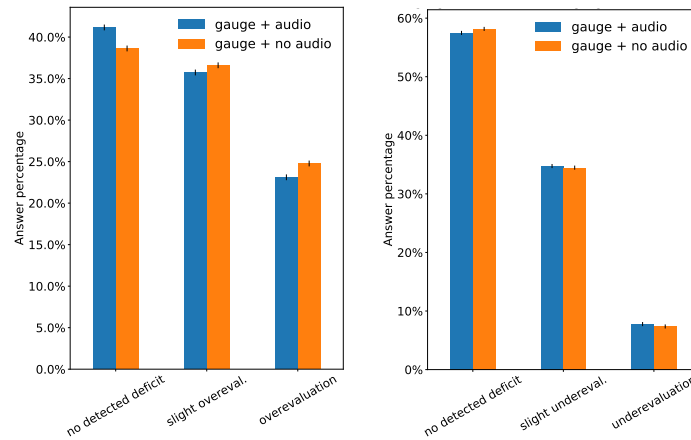


Fig. 6: Self-evaluation deficit distribution depending on audio recording when there is a gauge for poor (left) and excellent (right) performance answers with confidence interval at 95%. For the poor performance answers,  $N(\text{gauge} + \text{audio})=17,698$ ,  $N(\text{gauge} + \text{no audio})=18,350$ . For the excellent performance answers,  $N(\text{gauge} + \text{audio})=68,068$ ,  $N(\text{gauge} + \text{no audio})=68,648$ .

exercise ( $d = 0.05$ ,  $p < 0.001$ ). There was a significant impact on *increasing* the number of student underevaluating when their performance was excellent ( $d = 0.03$ ,  $p < 0.01$ ).

Overall we can therefore answer positively to RQ1 as the provided scaffoldings seem to have helped students in reducing their self-evaluation deficits.

## 5 Impact of audio remediation feedback on self-evaluation and self-efficacy deficits

### 5.1 Methods

To answer to our second research question, we designed four possible remediation feedback recording (cf. Table 2) to be played after the student displayed one of the four deficit tags from Table 1. Students who displayed a slight overevaluation or slight underevaluation did not receive any remediation feedback. We did the randomization so that half of the students would always get a remediation when they showed a self-evaluation or a self-efficacy deficit (remediation group) and half of the students would never get it (control group).

Table 2: Audio remediation recordings. See Table 1 to get the corresponding answers to the perceived and desired difficulty prompts.

deficit tag	audio recording
underevaluation	You said it was too hard, but you did great!
	It seems like this exercise was actually pretty easy for you!
overevaluation	You said it was too easy, but you made some mistakes. That's okay!
	Maybe this exercise was a bit too hard for you right now.
avoiding difficulty	You said it was too easy and that you wanted easier exercises.
	You can challenge yourself next time
	and ask for harder exercises. You're doing great!
seeking difficulty	You said it was too hard, and you're right, this was a tricky exercise. That's okay. You can keep trying.
	Instead of doing something harder,
	you can ask for an easier exercise to help you practice.

### 5.2 Results

**Impact of underevaluation remediation.** In order to measure the impact of the underevaluation remediation, we selected the traces with excellent performance for students in both the control group and the remediation group. In these traces, we selected students whose first trace with an excellent performance showed an underevaluation (see Table 1 for the definition of deficits). We then computed the answer deficit distribution on their next trace with an excellent performance where they got a self-evaluation assessment. The results for both groups are shown in Figure 7 (left). We remind that all students had shown an

underevaluation deficit on their first answer to the self-evaluation assessment. We notice that the number of students showing *again* an underevaluation is significantly smaller for students in the remediation group than for students in the control group ( $d = 0.17$ ,  $p < 0.001$ , two-tailed hypothesis). Conversely, there are significantly more students for whom we detect no deficit. There are also significantly more students for whom we detected a slight underevaluation, which corresponds to students for whom the deficit was partially addressed only (indeed, students with slight underevaluation did not receive any feedback, so there is no reason to expect a change otherwise).

**Impact of overevaluation remediation.** We did the same analysis to measure the impact of the overevaluation remediation. In these traces, we selected students in both groups whose first trace with a poor performance showed an overevaluation. We then computed the answer deficit distribution on their next trace with a poor performance where they got a self-evaluation assessment. The results for both groups are shown in Figure 7 (right). We remind that all students had shown an overevaluation deficit on their first answer to the self-evaluation assessment. We notice that the number of students showing *again* an overevaluation is significantly smaller for students in the remediation group than for students in the control group ( $d = 0.18$ ,  $p < 0.001$ , two-tailed hypothesis). Conversely, there are significantly more students for whom we detect no deficit. Also there are significantly more students for whom we detect a “slight overevaluation” which means they had a poor performance but declared the difficulty of the exercise they got was “Just right”. We can interpret these students as students that may feel close to succeeding at this exercise, even though their current performance is not good yet. For example, a student that had 3 questions in the last exercise and got two wrong answers and then a good answer may feel the difficulty is “Just right” as their last answer was correct, though their performance is considered “poor”.

**Impact of avoiding difficulty remediation.** We conducted a similar analysis to measure the impact of the avoiding difficulty remediation. We selected in both groups the traces with an excellent performance and in these traces, we selected students whose first trace with an excellent performance showed they wanted to avoid difficulty. We then computed the answer deficit distribution on their next trace with an excellent performance where they got the SRL assessments. The results are shown in Figure 8 (left). We can notice that the number of students detected as wanting to avoid difficulty (see Table 1 for the definition) decreases drastically the next time students are asked the SRL assessments in the remediation group, when compared to students in the control group ( $d = 0.30$ ,  $p < 0.001$ , two-tailed hypothesis). However, we also detect an increase in the number of students showing some underevaluation which will be addressed in the discussion section.

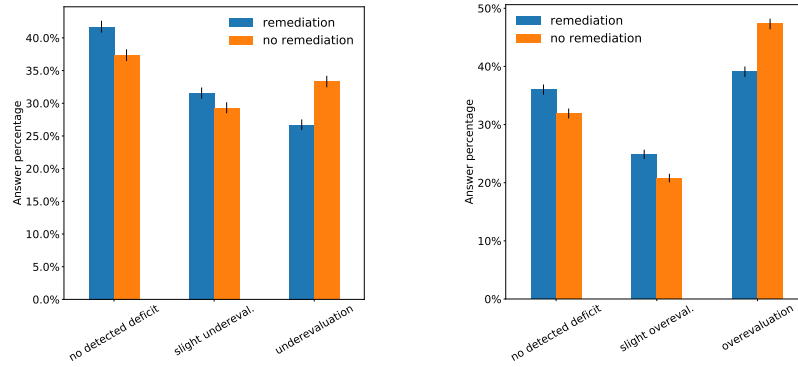


Fig. 7: **Left:** Self-evaluation answer distribution of the 2nd trace with excellent performance of a student when the 1<sup>st</sup> one was a *underevaluation* depending on remediation.  $N(\text{remediation}) = 11,397$ ,  $N(\text{no remediation}) = 11,196$   
**Right:** Self-evaluation answer distribution of the 2nd trace with poor performance of a student when the 1<sup>st</sup> one was an *overevaluation*, depending on remediation.  $N(\text{remediation}) = 11,513$ ,  $N(\text{no remediation}) = 11,476$ .

**Impact of seeking difficulty remediation.** Finally, we conducted a similar analysis to measure the impact of the seeking difficulty remediation. We selected in both groups the traces with a poor performance and in these traces, we selected students whose first trace with a poor performance showed they wanted to seek difficulty (see Table 1). We then computed the answer deficit distribution on their next trace with a poor performance where they got the SRL assessments. The results are shown in Figure 8 (right). We can observe that there is no significant difference between the answer distribution of the remediation group and the answer distribution of the control group ( $p > 0.05$ , two-tailed hypothesis), therefore we cannot conclude that the seeking difficulty remediation that we designed had any effect.

Overall, thanks to remediation there is a significant reduction in the number of students showing overevaluation or underevaluation. As we only do the analysis on the second trace with excellent (resp. poor) performance of a student and students were randomized into the two possible conditions, we can infer a causal relationship between the presence or not of the remediation and the difference in the answer distribution to the self-evaluation prompt. Moreover, as the self regulation prompts are only given with a probability  $\frac{1}{15}$  after finishing an exercise, the impact of one remediation is seen not immediately after it was given but later, suggesting lasting effects of remediation. We can therefore answer partially positively to RQ2.

## 6 Discussion and limits

In the first experiment, we noticed that the audio recording seemed to have a positive impact on students who had a poor performance and a negative impact

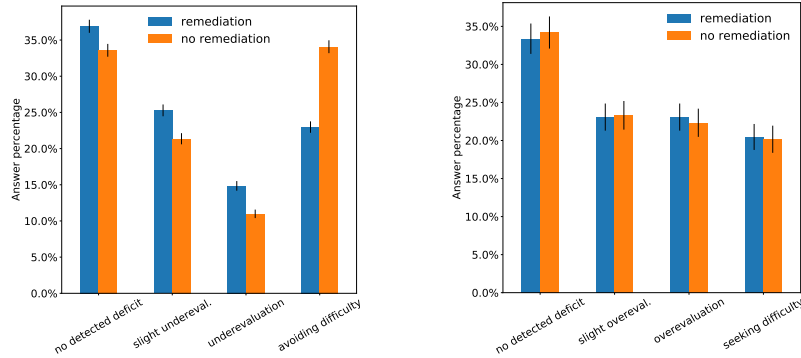


Fig. 8: **Left:** Self-evaluation answer distribution of the 2nd trace with excellent performance of a student when the 1<sup>st</sup> one was “avoiding difficulty”, depending on remediation.  $N(\text{remediation}) = 10,913$ ,  $N(\text{no remediation}) = 10,908$

**Right:** Self-evaluation answer distribution of the 2nd trace with poor performance of a student when the 1<sup>st</sup> one was “seeking difficulty”, depending on remediation.  $N(\text{remediation}) = 2,145$ ,  $N(\text{no remediation}) = 1,939$

on students who had an excellent performance. This indicates that the audio recording could allow a potentially distracted student to focus again after a poor performance. Conversely, it could distract a student who was focused after an excellent performance.

The most reliable way to assess SRL deficits is through direct questions to the students [2] and currently the frequency of the SRL statements is one out of fifteen. On the one hand, increasing the frequency of SRL remediation could improve SRL skills, however constant prompting can lead to an overall degraded perception of the learning environment [3]. Moreover SRL skills training should not come at the expense of literacy training which remains the main goal of the software. On the other hand, once we are able to detect that a student has good self-evaluation and self-efficacy skills, we could consider reducing the prompts, precisely so as not to waste time unnecessarily, paving the way for adaptive prompting.

The results of our second experiment showed that self-efficacy deficits (avoiding difficulty and seeking difficulty) are not ideally tackled with the audio remediation we designed. For the avoiding difficulty remediation, the number of students showing an underevaluation is higher in the remediation group than in the control group. Therefore, there should be a trade-off in the implementation of this remediation so that it does not impact students’ self-evaluation negatively. In other words, it appears that some students who were self-evaluating well but avoiding difficulty seem to resolve this contradiction not by asking harder exercise, but by declaring they did not feel like they were doing so well after all. This behavior could indicate either that they were not so sure in their self-evaluation in the first place, and that challenging their assessment made them hesitate, or that

they are somehow “gaming the system” by thinking that answering differently to the first question would prevent the system from raising the difficulty too much, which would be confirming the diagnosis of their difficulty avoidance behavior.

For the “seeking difficulty” deficit, we didn’t detect any significant effect. We can notice however that the sample size is smaller here than in previous analysis, so it is possible that with a similar number of samples, an effect would appear. Nonetheless there can also be explanations why this behavior is not as easy to impact as the other one, as “seeking difficulty” is a behavior consistent with a form of overconfidence, and discarding a system feedback is also consistent with overconfidence. For these students, a teacher intervention might be more appropriate. Alternative strategies could involve letting the student feel that they are incorrect by actually giving them a much harder exercise, or also asking them before an exercise how well they think they will succeed to confront their actual performance with their own self-evaluation *a priori* (and not only *a posteriori*).

Finally, we focused only on the first and second answers to students to the SRL prompts which are very local metrics. This allows us to measure precisely the impact of features on students’ answers. However, we did not describe the global SRL state of each student and its evolution over time after the first two SRL answers.

## 7 Conclusion and future works

In this work, we studied the answers to punctual self-evaluation and self-efficacy prompts of primary school (K-2) aged students on a literacy platform. We determined, using local metrics, the impact of visual and audio cues while answering self-evaluation prompts. The impact of visual cues (a gauge) is always significantly positive by decreasing the number of self-evaluation deficits. The impact of audio cues is mixed: it helps decrease the number of overevaluations but increases the number of underevaluations. Therefore it should be triggered only for poor performance students.

We also determined the effect of a remediation that was triggered when self-evaluation and self-efficacy deficits were detected. We managed to reduce significantly these deficits for some students with our actions locally (for self-evaluation more than for self-efficacy). Future works include the study of how remediation lasts in time or if it has to be reinforced regularly. We limited our scope to self-evaluation and self-efficacy in the SRL skills because we considered that these skills were measurable and could possibly be improved for students from Kindergarten to grade 2. Future works thus include the study of other SRL skills in primary school aged students.

This work underlines the possibility of scaffolding self-regulated learning skills in a web based application from a young age while learning another skill.

## References

1. Bannert, M., Reimann, P.: Supporting self-regulated hypermedia learning through prompts. *Instructional Science* **40**(1), 193–211 (Jan 2012).
2. Barnard, L., Lan, W.Y., To, Y.M., Paton, V.O., Lai, S.L.: Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education* **12**(1), 1–6 (Jan 2009).

3. Bouchet, F., Harley, J.M., Azevedo, R.: Evaluating Adaptive Pedagogical Agents' Prompting Strategies Effect on Students' Emotions. In: *Intelligent Tutoring Systems: 14th Int. Conf. LNCS*, vol. 10858, pp. 33–43. Springer, Montreal, QC, Canada (Jun 2018).
4. Dignath, C., Buettner, G., Langfeldt, H.P.: How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review* **3**(2), 101–129 (Jan 2008).
5. Hoffman, B., Spatariu, A.: The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology* **33**(4), 875–893 (Oct 2008).
6. Molenaar, I., Horvers, A., Dijkstra, R., Baker, R.S.: Personalized visualizations to promote young learners' SRL: the learning path app. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. pp. 330–339
7. Müller, N.M., Seufert, T.: Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction* **58**, 1–11 (Dec 2018).
8. Newton, P.M.: The Learning Styles Myth is Thriving in Higher Education. *Frontiers in Psychology* **6** (2015). , publisher: Frontiers
9. Panadero, E., Jonsson, A., Botella, J.: Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review* **22**, 74–98 (Nov 2017).
10. Schmitz, B., Wiese, B.S.: New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology* **31**(1), 64–96 (Jan 2006).
11. Schneider, W.: The Development of Metacognitive Knowledge in Children and Adolescents: Major Trends and Implications for Education. *Mind, Brain, and Education* **2**(3), 114–121 (2008).
12. Sergent, T., Daniel, M., Bouchet, F., Carron, T.: Assessing Children's Self-Regulation Deficits and Their Teachers' Perception: Towards a Co-Designed Remediation Through a Literacy Web Application. In: *Proceedings of ICALT 2021*. Online
13. Sergent, T., Daniel, M., Bouchet, F., Carron, T.: Predicting Young Students' Self-Evaluation Deficits Through Their Activity Traces. In: *International Conference on Educational Data Mining (EDM 21)*. Online
14. Zheng, L.: The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis. *Asia Pacific Education Review* **17**(2), 187–202 (Jun 2016).
15. Zimmerman, B.J.: Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal* **45**(1), 166–183 (Mar 2008).