



**HAL**  
open science

## Addressing Children's Self-Evaluation and Self-Efficacy Deficits in a Literacy Application

Thomas Sergent, Morgane Daniel, François Bouchet, Thibault Carron

► **To cite this version:**

Thomas Sergent, Morgane Daniel, François Bouchet, Thibault Carron. Addressing Children's Self-Evaluation and Self-Efficacy Deficits in a Literacy Application. 2021 International Conference on Advanced Learning Technologies (ICALT), Jul 2021, Tartu (online), Estonia. pp.141-143, 10.1109/ICALT52272.2021.00050 . hal-03353177

**HAL Id: hal-03353177**

**<https://hal.science/hal-03353177>**

Submitted on 23 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.

# Addressing Children’s Self-Evaluation and Self-Efficacy Deficits in a Literacy Application

Thomas Sergent  
Sorbonne Université, CNRS, LIP6 — Lalilo  
F-75005 Paris, France  
thomas.sergent@lip6.fr

Morgane Daniel  
Lalilo  
Paris, France  
morgane@lalilo.com

François Bouchet, Thibault Carron  
Sorbonne Université, CNRS, LIP6  
F-75005 Paris, France  
{francois.bouchet, thibault.carron}@lip6.fr

**Abstract**—The ability to self-regulate one’s learning (SRL) is considered to have a significant impact on educational outcomes. We present here a research work aiming first at detecting self-evaluation and self-efficacy deficits for young (5-7 years old) students, in the context of a literacy web application. From SRL answers we were able to characterize some answer patterns associated to SRL deficits, showing that around 30% of students seem to suffer from at least one of the four deficits considered in this study. We also surveyed close to 300 teachers to find out how they would like to be informed about their students’ SRL deficits and how they would address them, so that the remediation of deficits in the application could be co-designed with them.

**Index Terms**—self-regulated learning, primary school, co-design, educational data mining, self-evaluation, self-efficacy

## I. INTRODUCTION

According to Zimmerman [1], SRL is a three-phase cycle that repeats itself with each new task the learner is confronted with. In computer-based learning, SRL can be supported in various ways like prompts [2], [3] or automated feedback [4].

Improving children’s self-regulated learning (SRL) skills is seen as 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 [1]. In computer-based learning, SRL skills are generally viewed more as tools to support learning than as skills to be trained per se. Thus the measure that is used is the improvement of learning having support from these tools rather than an improvement in the SRL skills [5]. In their seminal work, Molenaar et al. [6] aimed at training the SRL skills of 5<sup>th</sup> grade students via dashboards: it therefore seems possible to measure and improve some SRL skills in relatively young children using computer-based systems. The differences with our work are that we have even younger students (5-7 years old vs. 9-10 years old for [6]) and that our metric is based directly on the student’s response to SRL questions and not on learning curves. From the various skills comprising SRL, we chose to focus on self-evaluation and self-efficacy. These two skills have been shown to be correlated with academic performance [7] and seemed to be isolated and independently measurable in young students. In this article, a first aspect we want to investigate is therefore the possibility of assessing self-evaluation and self-efficacy skills in young students.

A second aspect we want to investigate is relative to teachers’ awareness of the prevalence of some SRL deficits in their

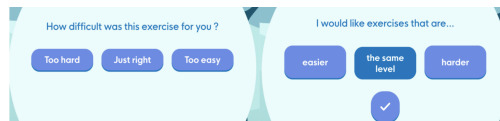


Fig. 1: Statement asking first the perceived difficulty (left) then the desired difficulty (right).

students and which approaches (if any) they currently use to try to reduce them. We therefore present as well the results from a survey to teachers using a literacy web applications in their classrooms to identify the most promising approaches in helping them to tackle their students’ SRL deficits.

More precisely, we will investigate the following research questions: (RQ1) Can we measure self-evaluation and self-efficacy abilities of young students learning with a web application? (RQ2) Are self-evaluation and self-efficacy prevalent issues for young students learning how to read? (RQ3) Are teachers aware of their students’ deficits and do they have ideas on how to address them?

## II. ASSESSING STUDENTS’ SELF-EVALUATION AND SELF-EFFICACY

### A. Context

Lalilo is one of the many web applications used by teachers in the classroom to help them implement a differentiated pedagogy. It is currently used by 40,000 English and French speaking kindergarten and elementary classes every week to strengthen literacy through series of exercises adapted to the students’ level, while providing the teacher with a dashboard to evaluate the students’ activities and progress. A typical session lasts 20 minutes with the student performing around 15 short exercises with 3 to 7 questions each. Student activities (e.g. logging in, time spent on an question/exercise, mistakes) are traced.

### B. Methods

1) *Data collection, filtering and cleaning*: To assess some aspects of students’ SRL skills, we introduced two statements (see Figure 1) which are randomly shown successively at the end of every fifteen exercises. The perceived difficulty statement aims at measuring the **self-evaluation** ability of the students.

TABLE I: Trace triplets tagged as having a deficit

Actual perf.	Perc. diff.	Desired diff.	Deficit
excellent	too hard	easier/same/harder	underevaluation
poor	too easy	easier/same/harder	overevaluation
excellent	too easy	easier/same	avoiding diff.
poor	too hard	harder/same	seeking diff.

The desired difficulty statement aims at measuring their **self-efficacy**, i.e. how they would react to their representation of the difficulty.

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 August 1<sup>st</sup> and December 11<sup>th</sup> 2020 on the Lalilo platform. We kept only the traces for which students had answered SRL statements.

We tried to limit potential sources of bias in our data by (1) filtering out students who had answered to less than 12 SRL statements, (2) removing students that seemed to answer randomly. There were two potential incoherent SRL answers combinations: evaluating the exercise as “too easy” (resp. “too hard”) but wanting “easier” (resp. “harder”) exercises. We considered that a student is potentially answering randomly if **both** incoherent SRL answers combinations appear in their traces. In the end, we had 556,357 traces from 28,263 EN students and 201,224 traces from 11,964 FR students.

### C. Deficit characterization

1) *At the trace level:* As a trace registers answers to each exercise question as well as to the two SRL assessments, 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 three values: excellent (all answers correct), poor (34% or less of the answers correct), and medium (for the remaining cases). Our goal is to compare the actual performance of a student with the difficulty they perceived - which is a cognitive representation - and then with the difficulty they desire for the next exercises. Indeed, comparing the desired difficulty to the actual performance may not be relevant if the students are biased in their perception of their real performance or in the difficulty of the task. From the performance, the perceived difficulty and the desired difficulty, we generate the so-called **trace deficit tag** displayed in Table I. We believe it is first necessary to solve potential self-evaluation deficits before addressing self-efficacy ones. Thus, if the answer to the perceived difficulty statement shows a self-evaluation deficit, then the trace deficit is this self-evaluation deficit. If there is no self-evaluation deficit but the answer to the desired difficulty statement shows a self-efficacy deficit, then two new tags are introduced: “avoiding difficulty” and “seeking difficulty”.

2) *At the student level:* Having tags at the trace level, our goal is then to detect some patterns in the student’s answers to globally characterize their SRL profile. Knowing for a student their number of SRL deficit tags (with Table I) and their number of “poor” and “excellent” performances, their **deficitRatio** is computed as the ratio between the number of answers tagged

TABLE II: Overall percentage of EN and FR students with one or more deficit (a student can only belong to one line).

Deficit(s)	EN % (N=28,263)	FR % (N=11,964)
no detected deficit	71.4%	63.7%
overevaluation	10.1%	11.5%
avoiding diff.	7.8%	11.4%
underevaluation	5.1%	5.1%
avoiding diff. - overeval.	3.3%	4.0%
seeking diff.	1.4%	2.9%
seeking diff. - undereval.	0.4%	0.6%
<i>incoherent combinations*</i>	0.5%	0.8%

\*6 additional lines with incoherent combinations are not represented here

as having a deficit and the number of related performances. Indeed, “overevaluation” and “seeking difficulty” are linked to poor performance while “underevaluation” and “avoiding difficulty” are linked to excellent performance. We chose the thresholds for the **deficitRatio** to be significant at 50% with at least 2 occurrences of the deficit. These values exclude random choice (which is at 33%) and a one time misclick by students.

### D. Results and discussion

Table II summarizes our findings. The three most frequent deficits are included in the four deficits defined in section II-C2: overevaluation, avoiding difficulty and underevaluation. The fourth (resp sixth) most frequently detected pattern refers to students having both “avoiding difficulty” (resp. “seeking difficulty”) and “overevaluation” (resp. “underevaluation”) deficits. Both are associated to a “too easy” (“too hard”) perceived difficulty so we hypothesize that students tagged with these two deficits do not properly understand the SRL statements and always click on “too easy” (resp. “too hard”). The order of deficit prevalence is similar in FR and EN students learning on Lalilo. These results show the possibility to reliably detect some SRL deficits in young students, and the frequencies we found suggest it would be necessary to address them. The goal of the following section is to confront these results with teachers’ perception.

## III. TEACHERS’ SRL KNOWLEDGE, REMEDIATION TECHNIQUES AND NEED FOR ASSISTANCE

### A. Methods

1) *Data collection:* We sent a survey to all Lalilo users teaching in French-speaking schools (SRL questions are also in the French version of Lalilo)<sup>1</sup>. We presented teachers the SRL statements embedded in Lalilo and the four SRL deficits we envisioned. Our goals were to (i) estimate teachers’ familiarity with SRL, their perception of deficit frequency and their practice of SRL training in their classrooms; (ii) know their interest in being informed of their students’ SRL deficits; (iii) understand how they would address deficits; (iv) know if they would accept a software intervention if it detects a SRL deficit.

Here we focus on (iii) only. We collected 298 answers to single-choice, multiple-choice and open-ended questions in October 2020. Open-ended answers were analysed manually, listing the suggestions contained in each answer. The

<sup>1</sup><https://cutt.ly/ubKvehJ>

suggestions in the result tables were determined *a posteriori*, depending on the answers given by teachers in our sample.

## B. Results

For the four categories of deficits considered (over- and under-evaluation, avoiding and seeking difficulty), 90% of the answers could easily be associated and tagged with a relatively small set of five to eight suggestions. Digging into the suggestions, we noticed some of them appeared in multiple categories of deficits:

- when performance is excellent (underevaluating or avoiding difficulty), with suggestions including: (1) giving harder exercises, (2) explaining students that one progresses mainly through confrontation with difficulty, (3) encouraging students to aim higher, (4) encouraging students to play a tutoring role.
- when performance is poor (overevaluating or seeking difficulty), suggesting to give the same or easier exercises.
- self-evaluation deficits (over/underevaluating), suggesting to have students explain their self-evaluation process.
- self-evaluation or poor performance deficits (over/underevaluating or seeking difficulty), suggesting that students count or visualize correct and incorrect answers. The other suggestions were specific to the targeted deficit.

## IV. DISCUSSION, LIMITS AND PERSPECTIVES

### A. Student deficit detection

Our first goal was to check the ability of a software to detect SRL deficits. We tried to detect students who responded randomly in our pre-analysis filters but some unexpected deficits (repetition of “too easy”/“too hard”) suggest a need for additional analyses before labeling students if an automatic intervention is then triggered based on this labeling.

Another potential limit is related to the threshold values for student’s tagging, as different threshold values impacted the absolute value of deficits although their relative frequency remained the same (results not shown here). Precise threshold tailoring will be needed to guarantee the sensibility and the specificity of detected deficits, as well as tagging only the most salient deficits not to overload the teacher’s dashboard.

### B. Teachers’ visions of students’ SRL deficits

Our second goal was to provide teachers with solutions that were not only supported by the scientific literature, but were consistent with what teachers do in their classrooms on a daily basis. In our study, we only used the answers of French and Quebec teachers, who may have a different approach and training than American teachers towards SRL.

We chose to ask open-ended questions not to restrict teachers’ suggestions or influencing them, which is the kind of approach that usually needs a second study to check whether another sample of teachers agrees and how they rank these suggestions.

Finally the main perspective is to implement teachers’ recommendations. Most encouragements can be addressed through gamification of the student interface (badges, diplomas or awards). Explanations (a form of formative feedback) would likely be provided through an audio file or a worked-out

example describing the expected behavior. Answers to SRL statements can also become an input for the adaptive learning algorithm to take into account teachers suggestions related to the difficulty of exercise to give afterwards. Having students count or visualize good and wrong answers could be achieved with a students-facing dashboard, an approach commonly used with Open-Learner Models for older learners [8]. Lastly, some suggestions are not easily transferable to an existing software such as encouraging students to play a tutoring role or having students explain their self-evaluation process to an adult.

## V. CONCLUSION AND FUTURE WORKS

To answer RQ1 and RQ2, we have presented a method for the detection of potential SRL deficits of students aged 5-7, asking about their perceived difficulty and desired difficulty for the following exercises. Analysing and characterizing answers patterns of around 40,000 students, we detected deficits in one third of the students, which is aligned with teachers’ perception of SRL deficits prevalence in their class. To answer RQ3, we collected suggestions of around 300 teachers on how they would tackle SRL deficits and highlighted similarities and differences in remediation of the deficits: a combination of actions to be taken by the student, explanations and encouragements. Most of the suggestions seem achievable by an e-learning software.

Future work includes comparing data on detected deficits with teachers’ perceptions of their students’ deficits to confirm the relevance of the detected deficits, then implementing remediations using the teachers’ suggestions that are transferable to a software, and finally evaluating their impact by studying temporal evolution of deficits with or without them.

## REFERENCES

- [1] B. J. Zimmerman, “Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects,” *American Educational Research Journal*, vol. 45, no. 1, pp. 166–183, Mar. 2008.
- [2] T. Lehmann, I. Hähnlein, and D. Ifenthaler, “Cognitive, metacognitive and motivational perspectives on reflection in self-regulated online learning,” *Computers in Human Behavior*, vol. 32, pp. 313–323, Mar. 2014.
- [3] F. Bouchet, J. M. Harley, and R. Azevedo, “Can Adaptive Pedagogical Agents’ Prompting Strategies Improve Students’ Learning and Self-Regulation?” in *Intelligent Tutoring Systems*, ser. Lecture Notes in Computer Science, A. Micarelli, J. Stamper, and K. Panourgia, Eds. Cham: Springer International Publishing, 2016, pp. 368–374.
- [4] A. T. Bimba, N. Idris, A. Al-Hunaiyyan, R. B. Mahmud, and N. L. B. M. Shuib, “Adaptive feedback in computer-based learning environments: a review,” *Adaptive Behavior*, vol. 25, no. 5, pp. 217–234, Oct. 2017.
- [5] L. Zheng, “The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: a meta-analysis,” *Asia Pacific Education Review*, vol. 17, no. 2, pp. 187–202, Jun. 2016.
- [6] I. Molenaar, A. Horvers, R. Dijkstra, and R. S. Baker, “Personalized visualizations to promote young learners’ SRL: the learning path app,” in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 2020, pp. 330–339.
- [7] S. E. Motlagh, K. Amrai, M. J. Yazdani, H. a. Abderahim, and H. Souiri, “The relationship between self-efficacy and academic achievement in high school students,” *Procedia - Social and Behavioral Sciences*, vol. 15, pp. 765–768, 2011.
- [8] R. Bodily, J. Kay, V. Alevan, I. Jivet, D. Davis, F. Khakaj, and K. Verbert, “Open learner models and learning analytics dashboards: a systematic review,” in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, ser. LAK ’18. New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 41–50.