Luciole: The Design Story of a Serious Game for Learning English Created for (and with) Children

Anne-Laure Guinet[1][a], Audrey Serna[1][b], Léo Vanbervliet[1], Emilie Magnat[2][c], Marie-Pierre Jouannaud[4][d], Coralie Payre-Ficout[3][e] and Mathieu Loiseau[1][f]

[1]*INSA Lyon, CNRS, Universite Claude Bernard Lyon 1, LIRIS, UMR 5205, 69621 Villeurbanne, France*

[2]*Ecole Normale Supérieure de Lyon, CNRS, ICAR, UMR 5191, France*

[3]*Université Grenoble Alpes, LIDILEM, France*

[4]*Université Paris 8, TRANSCRIT, France*

Keywords: Serious Games, User-Centered Design, English as a Foreign Language (EFL), Young Learners.

Abstract: In France, elementary school curriculum mandates 54 hours of foreign language instruction for pupils, typically English. However, many teachers feel linguistically insecure. Addressing this, we propose Luciole, a serious game designed to introduce 6 to 8-year-olds to English. Focused on oral comprehension, Luciole allows for autonomous use, facilitating differentiated activities. Luciole was developed through an iterative and inclusive series of design-development-user test-feedback analysis-redesign. Throughout this process, various methods were employed, including participatory workshops with children in the context of real classroom sessions. This article, through the description of the various design stages and successive prototypes, seeks to clarify a number of more general design issues related to serious learning games intended for use in the classroom, by pupils new to English, independently, and under the supervision of their teacher, who may be unfamiliar with the use of digital tools for learning.

1 INTRODUCTION

In 2002, a major change was orchestrated by the French government in terms of status of Foreign Language (FL) learning in elementary schools (Dat and Spanghero-Gaillard, 2005). The *Official Bulletin (BO)*[1] of February 14 2002 made FL an official subject of elementary school with a weekly hourly rate (MEN/MR, 2002). This decision was turned into a law in 2013 (MEN, 2013).

Despite this intention, the enforcement of such policies is not trivial. Elementary school teachers need to achieve equilibrium between "fundamental" skills and other mandatory subjects when some are integrated to the national evaluation campaigns and some are not. Often, FL play the part of adjustment variable, not only because of the difficult balance between subjects but also due to the teachers' own difficulties with FL. Few elementary school teachers are FL specialists and many feel linguistically insecure (MENJ, 2019; Delasalle, 2008). One of their difficulties is the priority set on the oral language, going from comprehension to production through repetition (MEN, 2015, p. 29). Some teachers fear being unable to produce acceptable utterances and fail to familiarize pupils with appropriate phonological references (Delasalle, 2008).

We thus decided to create a Serious Game (SG) for young learners of English (6–8 year-olds). We wanted to harness the possibilities offered by games to provide teachers with an application that learners could play autonomously, thus providing opportunities for differentiated teaching, while also providing learners with access to some of the building blocks of subsequent English learning activities (input and lexicon, culture, comprehension strategies). The product of this work is called "Luciole"[2].

[a] https://orcid.org/0000-0002-2957-1849
[b] https://orcid.org/0000-0003-1468-9761
[c] https://orcid.org/0000-0001-8857-9405
[d] https://orcid.org/0000-0003-1036-381X
[e] https://orcid.org/0000-0002-0156-9122
[f] https://orcid.org/0000-0002-9908-0770

[1]The French education ministry enforces decisions through publishing Bulletin Officiel (BO).

[2]*LUdique au service de la CompréhensIon Orale en Langue Étrangère i.e.* playing for improving oral comprehension of a foreign language

Guinet, A.-L., Serna, A., Vanbervliet, L., Magnat, E., Jouannaud, M.-P., Payre-Ficout, C. and Loiseau, M.
Luciole: The Design Story of a Serious Game for Learning English Created for (and with) Children.
DOI: 10.5220/0013285900003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 1*, pages 380-391
ISBN: 978-989-758-746-7; ISSN: 2184-5026

This paper is a case study of a long-term design process in a real world setting involving several hundreds of end users (teachers and pupils). It explains the rationale behind various design iterations carried out over several years of development. We believe that our approach highlights key considerations relevant to designing and testing learning games for children. The main research questions addressed here are:

1. How to design an experimental protocol to test a SG "in the wild"? And what are its consequences in the design of the application?
2. How to identify areas for improvement in an application that already demonstrated positive outcomes in terms of knowledge acquisition?
3. How to effectively engage young children in the design process, considering their limited familiarity both with tablet use and with SG?

To tackle these questions, we propose a "design story". It begins with a scientific context closely related to 1, followed by its influence on the first versions. Then it tells the part played by user-centered design methodology for education in the subsequent iterations and addresses research questions 2 & 3. Throughout this process, we extract design requirements (DR), identified [DR$_*$], that drove this design and can be applied to other projects.

2 DESIGNING A LANGUAGE LEARNING GAME FOR 6–8 YEAR-OLDS

Luciole was developed over the course of 8 years within different projects and fundings. Its integration into those projects represented a set of both opportunities and constraints, which impacted greatly the design process of the game.

Since 2016, several versions of Luciole have been developed. Each version has followed an iterative cycle: 1. design (within the team or integrating stakeholders in collaborative design sessions); 2. development; 3. user tests; 4. data collection and analysis.

Each cycle builds on the previous iteration, integrating users' feedback from the previous cycle. The data collection methods can vary from one iteration to the other according to its objectives, and the time and design constraints. In this article, we present the various methods for gathering needs, developing and testing prototypes, and analyzing results used over the project. Our main research questions for this paper being methodological, we describe more extensively the processes linked with the experimental protocol, the identification of areas for improvement and with the integration of young children in the process.

2.1 On the Design Constraints of Experimental Design

Luciole finds its origin in the Fluence project (2017-2022). It aimed to design and validate digital tools supporting learning (Mandin et al., 2021). Three applications were developed, two targeting reading skills (EVAsion & Elargir) and Luciole supporting English as a Foreign Language (EFL). To validate the applications, we needed to prove that each one improved the skill-set it targeted. To do so, the longitudinal protocol (Mandin et al., 2021; Loiseau et al., 2024) integrated 700+ pupils of 37 schools (fig. 1). They were separated in groups of consistent distribution of rural vs. urban schools, socially advantaged vs. disadvantaged schools. To ensure that control and experimental groups were comparable, they were created with a consistent distribution of pretest scores.

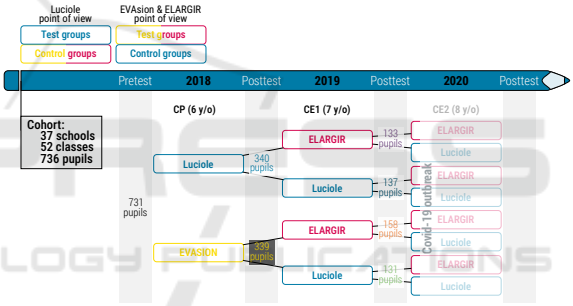


Figure 1: Fluence experimental design.

In a "design" paper, it is fitting to detail the implications of the experimental design. The tools were paired to neutralize the Hawthorne effect, which states: "those who perceive themselves as members of an experimental or otherwise favored group tend to outperform control groups, even in the absence of applied variables" (Koch et al., 2018, p. 3). In other words, the test group for each application is the control group for the other one. This creates constraints: [DR$_1$] Skill-sets targeted should not overlap across tested applications. To fulfill this Design Requirement (DR), there needs to be deep understanding of the skill-set targeted by the other team(s).

In our case, the core hypothesis behind EVAsion (a reading app) is that "optimal intervention for Visual Attention Span (VAS) enhancement should be based on the properties of action video games while requiring parallel processing of targets that gradually increased in number of visual elements" (Valdois et al., 2024, p. 3). In our case, [DR$_1$] translated into "the game should not use action video game structures

(Green et al., 2010, p. 203) or integrate written text".

The neutralization of the Hawthorne effect can result in situations where, in addition to one's own research questions, the protocol might apply constraints on the nature of the tools one can design.

2.2 On the Design of Serious Games

The first iteration began with a design phase that resulted in a *product backlog*, which allowed to recruit developers, designers and graphic artists[3]. The subsequent development phase was carried out using Scrum (Kniberg, 2007), allowing design tasks to occur during development (fig. 2).

2.2.1 On the Notion of Serious Game (SG)

The choice of creating a SG was directed by the properties attributed to them in literature: improve learner motivation (Garris et al., 2002; Reinhardt and Thorne, 2019), promote active learning (Vlachopoulos and Makri, 2017, p. 26), provide feedback on actions and assessment of player skills (Oblinger, 2004, p. 14), alter the perception of one's errors (Loiseau and Noûs, 2022, p. 69), not to mention take advantage of games design patterns which are adapted to learning (Gee, 2003). We define a serious game as "a game in the sense of (Duflo, 1997) designed, prescribed or used with both the aim of entertaining and an external finality targeting the player" (Loiseau and Noûs, 2022, p. 76). We will not engage in what Duflo means by game, by quoting this definition we want to separate from gamification which does not aim at producing a full-fledged game (Seaborn and Fels, 2015, pp. 14, 16, 27) but also recognize that we might fail in that endeavor and create an object which is not perceived as a game by the user. As a consequence, a DR emerges: [DR_2] the game must be perceived as a game by its players. We will see later on that it is not a tautology.

2.2.2 Narration in the Game

Among the design elements that are widespread in video games, the integration of a narrative (Domsch, 2013) was essential in our case but also constrained by DRs. [DR_3] A language learning game for 6-8 y/o should contain (almost) no written text is, in our case, highly consistent with [DR_1], as the other two Fluence applications target reading. Rather than being a consequence of [DR_1], [DR_3] stems from various factors. 6 year-olds are not yet readers; any written text might be understood extremely slowly if understood, thus impairing the engagement in the game.

[3]Over the project, we collaborated with two companies and 4 independent graphic artists and developers.

Second Language Acquisition (SLA) theory and official instructions also highlight this DR. The curriculum for our age target (6-8 year olds) makes oral language a priority (MEN, 2015, p. 31). Listening skills in particular, according to Krashen, constitute the foundation upon which other language skills are built. The aim of the second language classroom is to provide "comprehensible input" (Krashen, 1982, chap. 3), that is to say, to expose learners to utterances in their Second language (L2) that they can process. [DR_3] should thus be completed by [DR_4] A language learning game for 6-8 y/o should be the source of as much L2 input as possible. [DR_4] is also consistent with the situation of certain teachers we evoked in introduction. Since input is one of the building blocks of later language skills, a game introducing learners to EFL should provide opportunity for input by native speakers, thus discharging the teacher from some of the responsibility.

The narrative of the game addresses (at least in part) both DRs. The narrative integrates tasks in a coherent plot, but is also used to explain game structures, to introduce cultural elements about English speaking countries (MEN, 2015, p. 29) and to provide meta-linguistic information.

These DRs are at the core of the first versions of Luciole (cf. § 3), whose time frame did not allow us to integrate users in the original design. The subsequent design phases attempted to close that gap using user-centered design methodology.

2.3 User-Centered Design Methodology for Education

Involving end users in the development of learning environments has raised increasing interest due to its demonstrated benefits in enhancing usability, user experience, system adoption, and user engagement. Various methods and tools derived from Human-Computer Interaction (HCI) facilitate the integration of users in the design, implementation, and evaluation processes of interactive technologies. Because of developmental differences between children and adults, researchers have started to explore specific methods for the participation of children in design process, called Child-Computer Interaction (CCI). In their systematic literature review, Tsvyatkova and Storni identify methods, techniques and tools from User-Centered Design (Norman and Draper, 1986), participatory design (Simonsen and Robertson, 2013) and learner-centered design (Good and Robertson, 2006) developed or adapted to support children's involvement in design (Tsvyatkova and Storni, 2019). Their review indicates that most methods primarily support

design exploration and prototype evaluation, with a predominant focus on children aged 7 to 12. Prototyping methods are less frequently employed because they might be too difficult for children. Saiger *et al.* reached a similar conclusion in their review on children's involvement in game design, finding that children participated as true design partners in only half of the studies (Saiger et al., 2023). Five key factors were identified as influencing the effectiveness of children's involvement: comprehension, cohesion, confidence, accessibility, and time constraints. Most of the existing methods and tools focus on the evaluation of mock-ups or prototypes, especially with young children, to understand how they interact with the system and to identify their emotional responses. These tools are based on interviewing techniques — This-or-That (Zaman and Vanden Abeele, 2007), Contextual Laddering (Zaman and Vanden Abeele, 2010) — or propose instruments — Fun Toolkit (Read, 2008) — to identify children's engagement, likes or dislikes when interacting with products.

2.4 A Design Story

Luciole was created to respond to a need. Its design process was iterative and based on specific design requirements linked to its object and scientific context. We briefly evoked its main design methods, but an 8-year design process (Fig. 2) cannot be homogeneously summarized. In the next sections, we go through the different versions of the application.

Fig. 2 gives a color to each version of Luciole, showing its design and development phases as well as the user tests conducted and the analysis period. This figure illustrates the design cycle, both as a sequence of steps and in terms of how the previous phases provide input for the next. It also presents the various milestones to contextualize v5, for which we describe extensively the process which tackles research questions 2 and 3.

3 FLUENCE VERSIONS

3.1 v1: The Core

3.1.1 Game Structures

The design period of v1 corresponds to the definition of the core DRs and the foundations for all subsequent versions (cf. § 2). Once they were identified, we started to establish the game's curriculum. We separated the use of *constructions*[4] in three phases: 1. introduction of new *constructions*; 2. training on said constructions, used in simple utterances; 3. contextualization of trained constructions, used in more complex utterances, sometimes mixing constructions from diverse activities.

Despite its critiques, we adapted the Presentation-Practice-Production (PPP) model for its reliability for lower level learners (Anderson, 2016). Integrating these phases in the narrative aims to make every learning activity meaningful and targets the realization of an action oriented task. It thus confirms potentialities of games in language learning (Cornillie et al., 2012).

The narrative places the learner in the shoes of Sasha, a French child recruited by an intelligence agency to save animals. This choice owes to the special relationships children can develop with pets (Cassels et al., 2017). Sasha's missions take him to English speaking countries. He is helped by his mentor, Ash, who speaks French with a heavy English accent (to help the learner discover the sounds of English, and also to help focus on communicating and to play down possible production errors) and Hartguy, the solely English speaking coach.

To address [DR_2] through diversity in the player activity, we isolated 3 interaction matrices:

- identify & touch: The system provides an audio stimulus and the player needs to identify the sprite corresponding to the answer and touch it. It is mostly used in training and contextualization mini-games where one sprite is the correct answer for a given cue. In v1, Introduction phases always used an adaptation of this mini-game: players need to find the relevant sprites to hear the new vocabulary;
- drag 'n drop: The system provides an audio stimulus prompting the player to identify a draggable sprite among others and drop it on an associated drop zone. It is also used both in training and contextualization mini-games;
- remember: based on the metaludic rules (Silva Ochoa, 1999, p. 277) of memory games such as "Simon" or "Touch Me"[5], it is used for vocabulary memorization, and spans across introduction and practice.

We used these matrices to create 49 activities in v1. To provide feedback on the learner's progression in

[4]Strictly speaking *constructions* are "stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns" (Goldberg, 2003, p. 219). In this article, we also add phonemes though they do not have a function and should not be considered as such.

[5]https://en.wikipedia.org/wiki/Simon_(game)

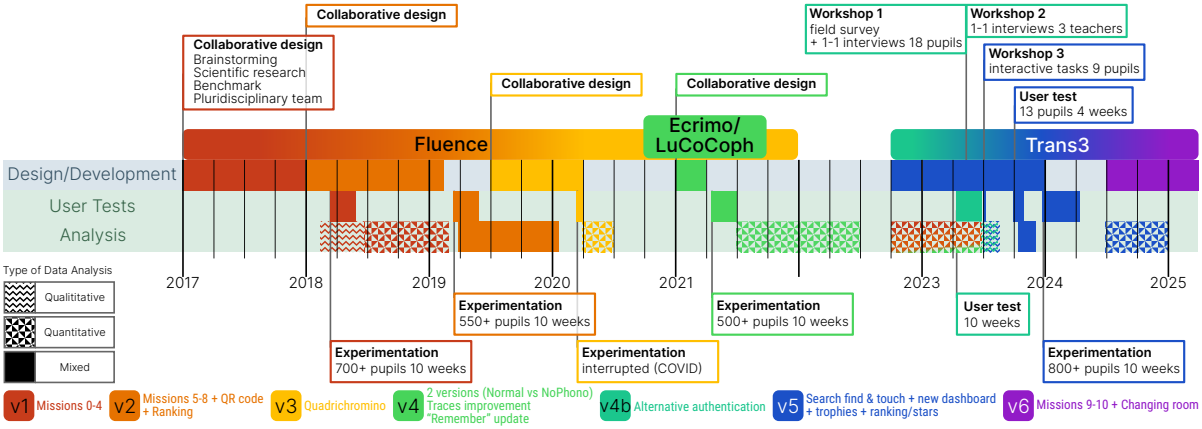
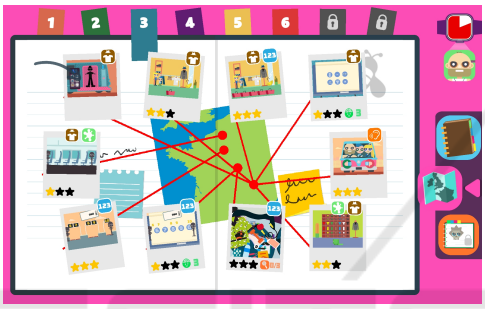
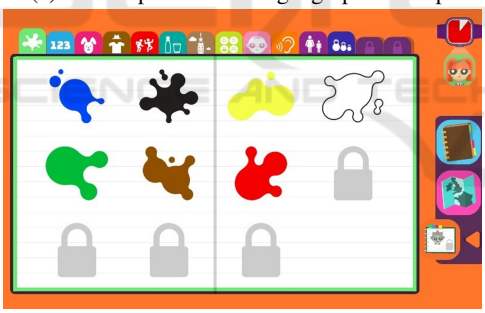


Figure 2: Luciole design story timeline (detailed version available online).



(a) The Map in v5 adds a geographical map.



(b) The notebook contains all the vocab learned.

Figure 3: Screenshots of the Luciole serious game.

the game we use a map: each activity is represented by a logo on the map and is associated with a score (0 to 3 stars) (Fig. 3a). Each activity unlocks the next one. Learners can replay activities from the map or the training center, where they are sorted by theme. To help learners keep track of what they have learned, we introduced a notebook displaying all the words they worked on, again grouped by theme (Fig. 3b).

3.1.2 Results

The game was tested for 10 weeks (prescription: 3×20 min/week). We collected pre/post-tests for 679 CP pupils (cf. fig. 1). We retrieved interaction traces for 310 out of 340 Luciole users and teacher feedback. Global results were positive: Luciole groups outperformed control group in the oral English comprehension tasks[6] (Mandin et al., 2021; Loiseau et al., 2024).

Teachers warned us that remember games discouraged pupils. It also appeared that some children advanced too fast in the game and then were lost for lack of mastery of previous content.

3.2 v2 & v3: Improving Reflexivity and Expanding the Game

3.2.1 Design Choices

Both qualitative feedback demonstrated lackluster realization of [DR$_2$]: such difficulty can disengage learners. We iterated over remember matrix. The second remark pointed to [DR$_5$] Advancing to the next mini-game should neither be too punitive nor too permissive. The quest for such balance led us to create a global score[7]. To motivate users, we associated milestones with a "spy rank" and created checkpoints that prevented players from accessing an activity unless they had reached a given rank.

A new interaction matrix was created to improve immersion in the game using non-digital resources (Loiseau et al., 2021). In the QR code mini-game, players scan one or more QR codes in response to oral instructions. QR codes are on a map of the British Isles. It concretizes Sasha's trip and diversifies game structures. At the end of each mission (some 10 activities), a new suspect is interrogated and the next sus-

[6] English score: $Luciole = 11.18$, $SD = 3.5$; $EVAsion = 8.41$, $SD = 3.33$; $t = -10.59$, $df = 677$, $p < 10^{-23}$.

[7] $Score_{global} = \sum_{A=1}^{n} max(Score_{activity}(A))$ with A the activity, 1 the 1st activity in the progression and n the furthest activity reached by the player.

pect is identified by scanning QR codes on the poster. v2 contained 94 mini-games.

v1 provided only usage information. In v2, we [DR$_6$] use the game to gather quantitative information from the learner's perception of the game and their learning. The narrative allows for self-assessment questionnaires every four missions. Sasha goes home with his mother who asks questions about his adventures. These questions constitute the self-assessment.

3.2.2 Results

The cohort of v1 was split in 4 subgroups. We collected pre/post-tests for 559 CE1 pupils distributed in 4 subgroups (Fig. 1). We successfully retrieved the interaction traces (including self-assessment) for 405 Luciole players and again collected teacher feedback. The overall results were also positive[8] (Mandin et al., 2021; Loiseau et al., 2024). Yet, the group which used Luciole two years in a row did not fare better than the other Luciole groups. This was later explained by a combination of rolling releases on our part during the experiment and of a lack of updates on the part of the teachers (Loiseau et al., 2024).

The trace system was not immune to bugs, some of the self-assessment was lost. But we could process the perceived level of mastery across lexical themes (not detailed here). In terms of user experience, the responses that stood out were the following: 62.7% reported understanding the story. 53% found the game easy, 43.7% found it moderate (sometimes easy, sometimes difficult), and 3.3% found it very difficult. Lastly, all pupils gave the maximum score to their overall appreciation of Luciole.

According to teachers, despite some improvement, remember still needed some work.

3.2.3 v3

The tests of v3 were rapidly interrupted due to the covid-19 outbreak. It provides a new mini-game meant to easily extend the game with utterance-image pairs. It is included as a bonus to players who complete the whole story.

3.2.4 After the Fluence Project

An unexpected result of the Fluence project was that the v1 Luciole group fared better in phonological awareness in their First language (L1) (French) than the EVAsion group[9] (Charles et al., 2025). Phonological awareness was included in the tests as a predictor of reading/writing skills (Valdois et al., 2024). At the end of the Fluence project, we tried to investigate the matter with a joint project (called ECRIMO2/LuCOCoPh, see Acknowledgements). In order to do so, we replicated the Fluence protocol, using ECRIMO (a dictation app) for our control group. Although we replicated the results in terms of English skills[10], phonological awareness results were disappointing, which we could attribute to the control task and our lack of means for individual testing of the pupils (Charles et al., 2025).

v4 was also an opportunity to improve Luciole's trace collection system (cf. § 3.2.2) and redesign the remember after v2 teacher feedback.This project was also an opportunity to reflect on our design processes and the need to implicate the users more deeply in the design phases to accelerate game improvements.

4 INVOLVING CHILDREN

The end of our multiple experimentations also created time to look more deeply into the game interaction traces. After four iterations, the most undesirable behavior highlighted was the lack of mini-game replay. Failure or limited star gain did not change that behavior. For example, in the remember game, the average scores range from 0.07 to 1.19 stars (out of 3). Yet, they are not replayed to gain more stars. The "training center" (meant to identify activities where score improvement was most likely) was underused: it was displayed an average 15 times per pupil. This lack of replay can be detrimental to acquisition. Hence, [DR$_7$] The game should encourage players to replay mini-games in which they did not fare well.

[DR$_7$] concerns the understanding of specific features (and global user experience). For these reasons, we decided to adopt a user-centered approach to involve children for the next design iteration. As underlined in section 2.3, methods to involve children are directed toward prototype evaluation. Using proven tools is especially relevant because children tend to seek to please adults during design workshops, express few spontaneous contradictions, and have difficulty making and expressing clear decisions. They can also be influenced by group effects, with a leader

[8] English score: $Luciole - * = 10.45$, $SD = 4.3$; $EVAsion - Elargir = 8.21$, $SD = 3.77$; $t = -6.1$, $df = 321$, $p < 10^{-8}$.

[9] Wilcoxon rank-sum tests for Phoneme deletion task: $W = 55280$, $p < .005$ & Phoneme segmentation: $W = 67044$, $p < .0005$

[10] Mann-Whitney U test Luciole$_{full}$ vs. ECRIMO: $U = 17015$, $p < 10^{-4}$; Luciole$_{NoPhono}$ vs. ECRIMO: $U = 19123$, $p < 0.005$

sometimes monopolizing the floor. This section details a methodology to work with 6 to 8 year-olds through 4 interventions with pupils and teachers. Testimonies in this section were translated from French.

4.1 Workshop 1

From March to June 2023, a user test of v4b (v4 with alternative authentication) was carried out (2 sessions a week). In June, a field survey was carried out with 18 elementary school pupils (CP) from three different classes. They were asked about their use of Luciole, and to demonstrate in-game access to certain features. These one-on-one interviews helped better understand some concerns (e.g. lack of replayability).

The main obstacles highlighted were: most children are unable to show their progress in the game (number of stars earned, secret agent rank); they are unable to explain in their own words the notion of stars; more critically, they do not understand how to access certain features (the "training center", or the ways to see their progression: the map to replay mini-games — fig. 3a — or the notebook — fig.3b).

The positive aspects identified were linked to key game elements: the storyline (*"it's good to free the animals"*, *"I love when we go on adventures"*); rewards (*"I like winning medals"*); learning new words related to the activities (*"learning colors"*, *"dragging clothes"*); or the visual environment (*"the images and sounds are beautiful"*).

Less appreciated aspects included activities perceived as too challenging, such as the remember game, and instances of lengthiness in dialogues (*"I don't like when they chat for too long"*). The vast majority of children were able to recount the story, or at least its main elements (characters, setting, goal). They were also able to spontaneously and instantly provide between one and a dozen English words (primarily related to animals, numbers, and colors — consistently with in-game auto-evaluations).

4.2 Workshop 2

In parallel, we carried out semi-structured interviews with a group of three teachers from the same user test. They were unanimously positive about their experience. They posited that children learned new words and both understood and enjoyed the story. They report general enthusiasm at the start of each play session, and a sort of bond with the main characters.

They observed that more advanced pupils sometimes went to help those experiencing more difficulties. The main concern was autonomy: some children had difficulty understanding and applying instructions during certain mini-games, and they were unable to complete certain tasks without teacher intervention. Then they did not progress in the game scenario, which led to frustration and discouragement.

Teachers confirmed the children's lack of understanding of some features: rank progression, time management, the means (and interest) of replaying mini-games. They also reported that the remember mini-games were not appreciated by all pupils. Some considered them too hard and insufficiently rewarding. Although traces analysis revealed a high level of use of the notebook (fig. 3b), the interviewed teachers expressed a strong desire to encourage students to listen more times to words in the game, but also to use these words in group activities outside the game (they were enthusiastic about our suggestion of a non-digital sticker album, for example).

4.3 Workshop 3

Based on outcomes of previous iterations, confirmed by workshops 1 & 2, one of our main goals was $[DR_7]$ (replayability). To achieve this, we undertook a complete overhaul of the dashboard and reward system (stars and ranks). The objective is to make it easier for pupils to visualize their progress and increase their motivation to replay mini-games in which they underachieved. In a user-centered approach, we decided to involve children in the design process, rather than depend on their feedback like in workshops 1 and 2. As we have explained in introduction to section 4, involving such young children is not straightforward, we thus resorted to the tools introduced in section 2.3.

4.3.1 Participants

In June 2023, 9 French primary school students participated in a workshop: 2 CP and 7 CE1 (5 girls and 4 boys). This group was part of a mixed level rural class. All participants had played Luciole during the user test. Children were divided into three groups. Each task was carried out with all three groups. For each group, two adults were involved.

4.3.2 Protocol

The method employed was inspired by the laddering technique, based on the Means-End theory (Zaman and Vanden Abeele, 2010). The name "laddering" comes from the idea of climbing a ladder, where each question serves as a rung leading to deeper insights. This systematic probing aims to achieve more comprehensive understanding of the subjects' attitudes, motivations, and decision-making processes. This

technique is valuable because it helps uncover motivations that individuals may not be fully aware of or able to articulate autonomously.We used the laddering method by following this presentation schema: First, we start with a simple question related to the children's experience with the game "What do you like about playing with Luciole?"; then we explore their motivations by asking "Why?" to uncover the motivations behind their preferences. For instance, if a child mentions enjoying the game because *"it's fun"*, we ask, *"Why is it fun for you?"*

It is important to frame questions in a way children this age can understand and relate to. Instead of asking questions about values, we focused on concrete aspects like favorite characters, game features, or challenges they enjoy. Moreover, children often express themselves through play and imagination. We have to encourage them to expand on their answers using their creativity. Understanding what motivates and engages pupils allows for targeted improvements that enhance their learning experience.

There were two experimenters per group of children: the "interactor" who engages with the children (e.g., asks questions, organizes speaking turns, etc.), and the "note-taker" who records as much information as possible, especially non-verbal behavior (audio was recorded), and assists in workshop set-up.

Experimenters had general instructions: Simple attire recommended (avoid intimidating clothing such as suits or ties); Allow children to get accustomed to the presence of adults (icebreaker), and avoid appearing rushed; Adopt an attitude of ignorance to stimulate the involvement and sharing of information by the child. Position themselves in a rapport if possible equivalent by mentioning ignorance on certain points, which they can respond to. Moreover, it is advisable to adopt the least adult-like position possible (Zaman and Vanden Abeele, 2010, p. 160).

4.4 Tasks (Fig. 4)

The workshop was split in 5 stages with a 15-minute break between the second and third task.

Icebreaker and Recall of Luciole's Story. The workshop starts with a presentation involving all participants and outlining its objectives and tasks. A Luciole video is shown to introduce the game and help children recall it.The goal is to create a common setting and make it clear to the pupils that they participate to the same tasks no matter which subgroup they are. Pupils are encouraged to engage in spontaneous remarks both to engage them, notice we listen to them

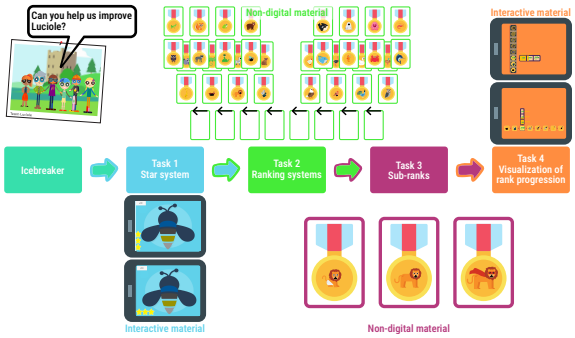


Figure 4: Workshop 3 task plan (Luciole v5).

but also to gauge their enthusiasm and to identify potential leaders who speak up frequently.

Task 1 — Star System. The aim is to improve the pupils' understanding of the stars they earn at the end of each mini-game. We created both paper visuals and interactive versions to chose between vertical and horizontal representation of the stars. We start by reminding pupils of the "this-or-that" instructions, then show the paper visuals (to avoid children rushing to the tablets — untimely clicks, ignoring instructions), while reminding them of how stars are earned in Luciole. After freely testing the two options on the tablet, the children indicate on their slate the number of the proposition they prefer. Their choice is the basis for a discussion where they explain their choices individually and collectively. The experimenter makes sure that all the children participate.

Task 2 — Ranks. In v4, children reported not knowing their "spy rank". The rank is supposed to motivate children and provide an indicator both of the progression in the game and of the level of mastery. In this task, the aim is to choose which of the four proposed animal rankings is more appealing, make sure children recognize each animal, and adjust the animal order within the ranking. Children write their preferences on their slates, then comment on each animal. The experimenter systematically asks questions to make children justify their choices (Why?).

Task 3 — Sub-Ranks. We created sub-ranks for each animal: it generates more opportunities to get a reward, to hear the animal name, and optimizes visualization on the tablet. Animals evolve from baby, to adult, to superhero. The task targets the children's understanding and acceptance of this progression. They sorted the stages of this evolution and justified their choices (Why?). Particular attention was paid to the "baby" rank: we asked each child whether they would accept being depicted as a baby lion wearing a diaper.

Task 4 — Visualization of Rank Progression. Previous tasks allow to gauge interest for various rank scales, but to address children not understanding that indicator, we need the representation of the scale to be clear. We created two mock-ups on the tablet, one vertical and one horizontal representation of the ranking system. Children could manipulate both modes on the tablet. Then each indicated their preferred visualization on their slate. The experimenter then had them explain their choices (Why?).

Material. Various material were necessary for this workshop. They were designed and created specifically for this study, but could be adapted and used for another study (Table 1).

Table 1: Materials used for the co-conception workshop.

Physical Material
Animal grade image cards proposed (32 cards) (4×8)
Ranking card track (divided into 2 A4 sheets) ×4
Selection number allowing to associate a number with a ranking or a prototype. A4 sheet divided into 4 or 1 sheet per number.
Adjective image cards associated with intra-grade progression (Lion — 3 cards Baby/Normal/Super)
Vertical example star presentation and horizontal to guide the child during this or that ×1
Vertical example gauge presentation and horizontal to guide the child during this or that ×1
Example sheet of video games using a star system
Numerical Material
Progress bar with grade scale (vertical/horizontal)
Star activity progress bar
Base - without star (i.e., to explain the activity)
Other Supports
Luciole video
Slates × 5
Whiteboard markers × 8
Camera or smartphone
Microphone and recording equipment (e.g., computer)

4.4.1 Results

Data collected came from audio recordings, verbatim, pictures (slates) and real-time completion of observation grids by adult observers.

Task 1. The consensus was in favor of the vertical view of the stars. The children's arguments were: climbing is vertical (*"we climb the stairs to go even higher"*); positive evolution goes up; more aesthetic and easier to understand. The horizontal view was considered unattractive and could hinder the completion of an activity for taking up more space on the screen. The children also indicated that they wanted star filling to be progressive (gauge effect).

Task 2. The order of preference for the animal rankings was: 1. **marine animals** (3 votes and 2 groups reaching consensus), 2. **birds** (2 votes and 1 group reaching consensus), 3. **food chain** (2 votes), 4. "cool" animals did not receive votes. The two favorite marine animals were, with two votes each, the swordfish and the dolphin. Numerous lively discussions led to adjustments, both of the ranking order (*"the octopus is stronger than the crab"*) and to the visual appearance of certain animals (*"the dolphin is too big"*). Grades not yet unlocked should be visible to whet players' appetite, but grayed out and padlocked to show that they are not yet earned.

Task 3. The sub-ranks were understood and unanimously approved (no reservations about the "baby" stage).

Task 4. Consensus was quickly reached on the horizontal view described as clearer and easier to understand, but many questions arose about the space it would take up on the screen.

4.4.2 Subsequent Development

The workshop was quickly followed by various developments : New graphics representing the marine animals: shrimp, seahorse, crab, octopus, seal, dolphin, swordfish, orca all declined in three sub-ranks each, with visual reinforcement of the notion of rank (medal) with the addition of a ribbon; Padlocking of not yet reached ranks; possibility to hear both locked and unlocked ranks; horizontal presentation of ranks, the sub-ranks are displayed vertically when the rank is touched.

4.5 Prototype Validation by Pupils

Traces analysis had shown that in introduction phases, children did not actually listen to the sounds they were supposed to become familiar with. They just touched the screen frenetically until all objectives were uncovered. We created a new mini-game (coined Search, find and touch), so that they could only hear one new item at a time to hear it clearly.

After developing the new ranking system and mini-game, we tested the new prototype with children who had never played the previous versions of Luciole. This study consists of three interventions, each one week apart. During each intervention, observers watch children participate in a Luciole game session, and then gather feedback and suggestions.

4.6 Outline of Interventions

Introduction and Game Overview. At the beginning of the first intervention, team members introduce themselves, present the Luciole project, and give a hint of the story-line, the characters and explain how to launch the game (account system).

Game Session and Observation. During each intervention, children play Luciole for 20 minutes, each time resuming their progression. One observer conducts shadowing observations (Czarniawska, 2007) based on a specific observational grid.

Qualitative Feedback. At the end of each 20-minute session, researchers conduct a semi-structured discussion with the children, teachers to gather feedback and suggestions for improvement.
Feature Testing Workshops. In the final two interventions, specific workshops are organized to test certain game features with small groups of pupils (N=4), aiming to collect targeted feedback.
Analysis and Team Member Brainstorming. After each intervention, the observations and workshop results are analyzed to identify and refine improvement areas for the next session.
Autonomous Session of Luciole in Class. Between each intervention, children continue their progress with their teacher during two autonomous in-class sessions.
Final Evaluation. The third intervention ends with pupils completing the System Usability Scale (SUS) (Vlachogianni and Tselios, 2022).

4.7 Results

Observation sessions and analysis of the verbatim showed a high level of understanding of the new playing instructions and rankings. Pupils unanimously approved the marine animal badges. They were proud to compare their progress with each other and share it with the teacher. The summary of comments made during the test workshops did not reveal any particular difficulties for pupils with the new Search, find and touch mini-game.

The mean score on the SUS questionnaire was 88.7 (standard deviation: 7.2), indicating excellent acceptability of the game (even with young subjects). The items with the highest scores concerned handling of the game, which was rated as easy to use, light and requiring no external assistance. Interest was also shown in using the game more frequently. Analysis of the 13 pupils' use of the Luciole game (with access to information such as progress, number of mini-games played, stars won, notes consulted, etc.) revealed that the time spent playing the game was as prescribed, and that the game's features were widely explored. The progress made by the pupils was remarkable.

5 DISCUSSION

This paper tells the story of the design of a serious game over eight years. Its design has been driven by DRs that either emanated from the learning objectives for the players or the experimental protocol itself. While the first source of constraints is common the second raises more questions. Should the design of a technological object be driven by how the object will be tested? In our case, the constraints of the protocol were in line with our scientific and learning objectives and the decision to join the project was straightforward. But this should not downplay this issue. Serious games have been pointed out for lack of longitudinal and/or empirical studies (Girard et al., 2013) and the neutralization of variables such as the Hawthorne effect come with constraints that should not be overlooked. In the v4 tests, the choice of the control group might be the cause for some disappointing results (Section 3.2.4).

Working within this set of constraints, the regular tests in ecological conditions allowed us to gather diverse data on the use of our game. We had regular interactions with teachers who signaled bugs and impressions on the use of the system, gathered interaction traces, collected observations, questionnaires and post-test data. This wide array of data both in nature and quantity allowed us to continuously improve our application. Qualitative information (for instance the class observation that children did not seem to pay attention in introduction phases of versions 1 through 4) could be verified thanks to quantitative data (traces analysis). In that, we did not only mobilize usage data to qualify the behavior of the learners but also to identify issues and improve the design.

Still, the fact that the game was created all along the Fluence project put pressure on the development process and let us settle in a rhythm of gathering information at the end of the tests (one iteration per year), instead of implicating the pupils and teachers directly. This rhythm showed its efficiency limits in the number of iterations necessary to fix our remember mini-game (cf. [DR_2]).

On the contrary, before the Trans3 project we had time to analyze our data and organize design workshops with actual end-users. Those proved very effective in the redesign of the visual research game or the ranking systems. To set-up these workshops we used design tools that were adapted to children (see § 4). Yet, we underestimated the length and frequency of digressions within those workshops. In turn, the time allotted to digressions made workshop longer than expected, and exhausted the children. Digressions should not be avoided, they participate in creating a trust environment, an atmosphere where children can really contribute. But they should be accounted for in the length of the design workshops.

We should also mention that shadowing is sometimes difficult to understand for a 6 year-old child who wants to advance in the scenario and asks for help (such interactions or lack thereof were very productive in the discussion part of the workshop).

Finally, it should be noted that the design tools we chose might not be adapted for other life cycles. Had we implicated the children when we created the game from scratch, we probably would not have relied on evaluation tools and prototype version choices. But the improvement of a tested and validated system was in line with our objectives.

6 CONCLUSION

Over the course of its design and development, Luciole has been tested on three cohorts of hundreds of children (twice on one cohort). In each experiment, the Luciole groups fared better than their counterparts regarding EFL. Each experiment took place under business as usual conditions, each group being given an application targeting separate skills to neutralize the Hawthorne effect. In one occasion, it also proved to have an effect not only on English but also on phonological awareness. Over the course of eight years of work we have made our design practices evolve to use various source of data.

Now that Luciole has reached a form of stability (an almost "final" version should be published in September 2025), many venues are open for experimenting. The influence of the context of use should be analyzed. Many research questions can be identified: how the learning context (school vs. formal out-of-school learning vs. informal learning) influences Luciole's outcomes; what in-class group activities to carry out to maximize Luciole's effect; or how and which non-digital material (and rewards) can influence children's engagement in the game, and their motivation for foreign languages, or for school.

ACKNOWLEDGEMENTS

The authors are grateful to the French government (France 2030) for its financial support through the Trans3 project (ANR-22-FRAN-0008), managed by the Agence Nationale de la Recherche (ANR). This work was also supported by

- the e-FRAN program ("e-FRAN Fluence" project, 2017–2022), funded by the "Programme d'Investissement d'Avenir" (PIA2) handled by the Caisse des Dépôts et Consignations (CDC);
- the ASLAN project (ANR-10-LABX-0081) of the Univ. de Lyon, for its financial support within the French program "Investments for the Future" operated by the ANR: funding of LuCOCoPh.
- the Pole Pilote Pégase, French government funded "Territoires d'innovation pédagogique" of the "Investments for the Future" program, operated by the CDC: funding of the ECRIMO2 project.

REFERENCES

Anderson, J. (2016). Why practice makes perfect sense: the past, present and future potential of the PPP paradigm in language teacher education. *English Language Teaching Education and Development (ELTED)*, 19(1):14–22.

Cassels, M. T., White, N., Gee, N., and Hughes, C. (2017). One of the family? Measuring young adolescents' relationships with pets and siblings. *Journal of Applied Developmental Psychology*, 49:12–20.

Charles, E., Magnat, E., Jouannaud, M.-P., Payre-Ficout, C., and Loiseau, M. (2025). An English listening comprehension learning game and its effect on phonological awareness. In Frisch, S. and Glaser, K., editors, *Early Language Learning in Instructed Contexts*, Language Learning & Language Teaching. John Benjamins, Amsterdam/Philadelphia.

Cornillie, F., Thorne, S. L., and Desmet, P. (2012). Digital games for language learning: from hype to insight? *ReCALL*, 24:243–256.

Czarniawska, B. (2007). *Shadowing: And other techniques for doing fieldwork in modern societies*. Liber.

Dat, M.-A. and Spanghero-Gaillard, N. (2005). L'enseignement des langues et cultures étrangères à l'école primaire : un exemple d'utilisation de document authentique multimédia. *Corela*, (HS-1). Number: HS-1 Publisher: Université de Poitiers.

Delasalle, D. (2008). Enseigner une langue à l'école : a-t-on les moyens de relever ce défi dans le contexte actuel ? *Études de Linguistique Appliquée*, 3(151):373–383.

Domsch, S. (2013). *Storyplaying: Agency and Narrative in Video Games*. De Gruyter.

Duflo, C. (1997). *Jouer et philosopher*. Pratiques théoriques. Presses Universitaires de France, Paris.

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4):441–467.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillan, Basingstoke. OCLC: 637002203.

Girard, C., Écalle, J., and Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, 29(3):207–219.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

Good, J. and Robertson, J. (2006). Carss: A framework for learner-centred design with children. *International Journal of Artificial Intelligence in Education, International Journal of Artificial Intelligence in Education*, 4(16):381–413.

Green, C. S., Li, R., and Bavelier, D. (2010). Perceptual learning during action video game playing. *Topics in Cognitive Science*, 2(2):202–216.

Kniberg, H. (2007). *Scrum and XP from the Trenches – How we do Scrum*. C4Media.

Koch, M., von Luck, K., Schwarzer, J., and Draheim, S. (2018). The novelty effect in large display deployments — experiences and lessons-learned for evaluating prototypes. In Gross, T. and Muller, M., editors, *Proceedings of 16th European Conference on Computer-Supported Cooperative Work — Exploratory Papers*. European Society for Socially Embedded Technologies (EUSSET).

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Language teaching methodology series. Pergamon Press, Oxford.

Loiseau, M., Jouannaud, M.-P., Payre-Ficout, C., Émilie Magnat, Serna, A., and Guinet, A.-L. (2024). Large scale experimentation of a serious game targeting oral comprehension in esl for 6–8 year-olds: Some lessons learnt. In *CALICO 2024 — Confluences and Connections: Bridging Industry and Academia in CALL*.

Loiseau, M. and Noûs, C. (2022). À la recherche de l'attitude ludique. In Silva, H., editor, *Regards sur le jeu en didactique des langues et des cultures*, volume 9 of *Champs Didactiques Plurilingues*, chapter 1.3, page 61–83. Peter Lang Verlag, Bruxelles, Belgique.

Loiseau, M., Payre-Ficout, C., and Émilie Magnat (2021). Le qr code dans le jeu sérieux "luciole" : lien entre le monde virtuel et le monde réel pour le développement de la compréhension orale en anglais à l'école. In *Actes de "Objets pour apprendre, objets à apprendre : Quelles pratiques enseignantes pour quels enjeux?"*.

Mandin, S., Zaher, A., Meyer, S., Loiseau, M., Bailly, G., Payre-Ficout, C., Diard, J., Fluence-Group, and Valdois, S. (2021). Expérimentation à grande échelle d'applications pour tablettes pour favoriser l'apprentissage de la lecture et de l'anglais. In *10e conférence EIAH — Transformations dans le domaine des EIAH : innovations technologiques et d'usage(s)*.

MEN (2013). L'enseignement des langues vivantes étrangères et régionales (articles 39 à 40).

MEN (2015). Programmes d'enseignement de l'école élémentaire et du collège. Bulletin officiel spécial du 26 novembre 2015 386, Ministère de l'Éducation Nationale, Paris.

MENJ (2019). Guide pour l'enseignement des langues vivantes. Technical report, Ministère de l'Éducation Nationale et de la Jeunesse, Paris.

MEN/MR (2002). Horaires et programmes d'enseignement de l'école primaire. Bulletin Officiel hors-série du 14 février 2002 1, Ministère de l'Éducation Nationale et Ministère de la Recherche, Paris.

Norman, D. and Draper, S. (1986). *User Centered System Design: new perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associate, London, UK.

Oblinger, D. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education*, 2004(1):1–16.

Read, J. C. (2008). Validating the fun toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, 10(2):119–128.

Reinhardt, J. and Thorne, S. L. (2019). *Digital Games as Language-Learning Environments*, chapter 17, page 409–435. MIT Press, Cambridge, MASS.

Saiger, M. J., Deterding, S., and Gega, L. (2023). Children and young people's involvement in designing applied games: Scoping review. *JMIR Serious Games*, 11:e42680.

Seaborn, K. and Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74:14–31.

Silva Ochoa, H. (1999). *Poétiques du jeu. La métaphore ludique dans la théorie et la critique littéraires françaises au XX^{e} siècle.* Thèse, Université Paris 3 — Sorbonne Nouvelle, Paris.

Simonsen, J. and Robertson, T. (2013). *Routledge International Handbook of Participatory Design*. Routledge, New York, USA.

Tsvyatkova, D. and Storni, C. (2019). A review of selected methods, techniques and tools in child–computer interaction (CCI) developed/adapted to support children's involvement in technology development. *International Journal of Child-Computer Interaction*, 22:100148.

Valdois, S., Zaher, A., Meyer, S., Diard, J., Mandin, S., and Bosse, M. L. (2024). Effectiveness of visual attention span training on learning to read and spell: A digital-game-based intervention in classrooms. *Reading Research Quarterly*.

Vlachogianni, P. and Tselios, N. (2022). Perceived usability evaluation of educational technology using the system usability scale (SUS): A systematic review. *Journal of Research on Technology in Education*, 54(3):392–409.

Vlachopoulos, D. and Makri, A. (2017). The effect of games and simulations on higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education*, 14(1).

Zaman, B. and Vanden Abeele, V. (2007). How to measure the likeability of tangible interaction with preschoolers. In *CHI Nederland*, page 57–59, Woerden. Infotec Nederland.

Zaman, B. and Vanden Abeele, V. (2010). Laddering with Young Children in User EXperience Evaluations: Theoretical Groundings and a Practical Case. In *Proceedings of the 9th International Conference on Interaction Design and Children*, IDC '10, page 156–165, New York, NY, USA. Association for Computing Machinery. event-place: Barcelona, Spain.