HAL open science

Conversational agents for fostering curiosity-driven learning in children

Rania Abdelghani, Pierre-Yves Oudeyer, Edith Law, Catherine de Vulpillières, Hélène Sauzéon

▶ **To cite this version:**

Rania Abdelghani, Pierre-Yves Oudeyer, Edith Law, Catherine de Vulpillières, Hélène Sauzéon. Conversational agents for fostering curiosity-driven learning in children. International Journal of Human-Computer Studies, 2022, 167, pp.102887. 10.1016/j.ijhcs.2022.102887 . hal-03779096

HAL Id: hal-03779096

https://hal.science/hal-03779096v1

Submitted on 1 Feb 2023

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.

arXiv:2204.03546v2 [cs.HC] 12 Apr 2022

Conversational agents for fostering curiosity-driven learning in children

Rania Abdelghani[1,3], Pierre-Yves Oudeyer[1], Edith Law[2], Catherine de Vulpillières[3], and Hélène Sauzéon[1]

[1]Flowers team, Inria Bordeaux, France
[2]HCI lab, Waterloo University, ON, Canada
[3]EvidenceB, Paris, France

April 13, 2022

Abstract

Curiosity is an important factor that favors independent and individualized learning in children. Research suggests that it is also a competence that can be fostered by training specific metacognitive skills and information-searching behaviors. In this light, we develop a conversational agent that helps children generate curiosity-driven questions, and encourages their use to lead autonomous explorations and gain new knowledge. The study was conducted with 51 primary school students who interacted with either a neutral agent or an "incentive" agent that helped curiosity-driven questioning by offering specific semantic cues. Results showed a significant increase in the number and the quality of the questions generated with the "incentive" agent. This interaction also resulted in longer explorations and stronger learning progress.

Together, our results suggest that the more our agent is able to train children's curiosity-related metacognitive skills, the better they can maintain their information-searching behaviors and the more new knowledge they are likely to acquire.

Keywords: Human-computer interface, Cooperative/collaborative learning, Teaching/learning strategies, Improving classroom teaching

1 Introduction

Teaching strategies that rely on giving direct instructions to children have long been used in our classrooms and proven to be rather effective with learning standardized pedagogical content. Yet it has been shown that children learn and memorize the information better when they look for it themselves [41]. Such curiosity-driven practices can also offer the individualized and independent learning environments that today's education systems need. Cranton [12] suggests that epistemic curiosity (i.e the desire for knowledge that motivates learning new material) can be fostered in classrooms by providing activating events and time for puzzlement and critical self-reflections. Jirout [41] also suggests that promoting curiosity comes from promoting comfort with uncertainties and encouragement to think about what one does not know, to ask questions and to explore one's environment. The importance of curiosity has gathered much attention in research, as it leads to a better learning experience and to the raising of the child's cognitive awareness of his own knowledge.

Indeed, research, such as in [5], shows that the brain is made to memorize by asking questions and not simply by receiving information through reading or listening to it. These studies show that students who ask questions and make hypotheses perform much better than those to whom the information was simply passed. On another hand, Berlyne [4] explains that being aware of one's own missing knowledge is the crucial factor that incites individuals to seek information. Taken together, these findings lead us to suggest that training the curiosity-driven information-seeking behaviors should start with training the metacognitive skill of identifying one's knowledge limits and needs.

In this work, we take inspiration from curiosity models and frameworks in order to design a a novel educational platform called "Kids Ask" involving a conversational agent that trains curiosity-based question-asking skills and encourages their use to lead autonomous explorations of their learning environments. We also investigate the role of these curious behaviors in helping children acquire new knowledge (see Figure1). Beyond exploring the use of conversational agents only to train curiosity-driven question-asking skills, our study goes one step further and investigates also how such agents can help children use this skill to move forward with their learning. It contributes both to the ongoing research on novel learning approaches in general and to our understanding of how to design effective interactive platforms for promoting curiosity-driven learning.

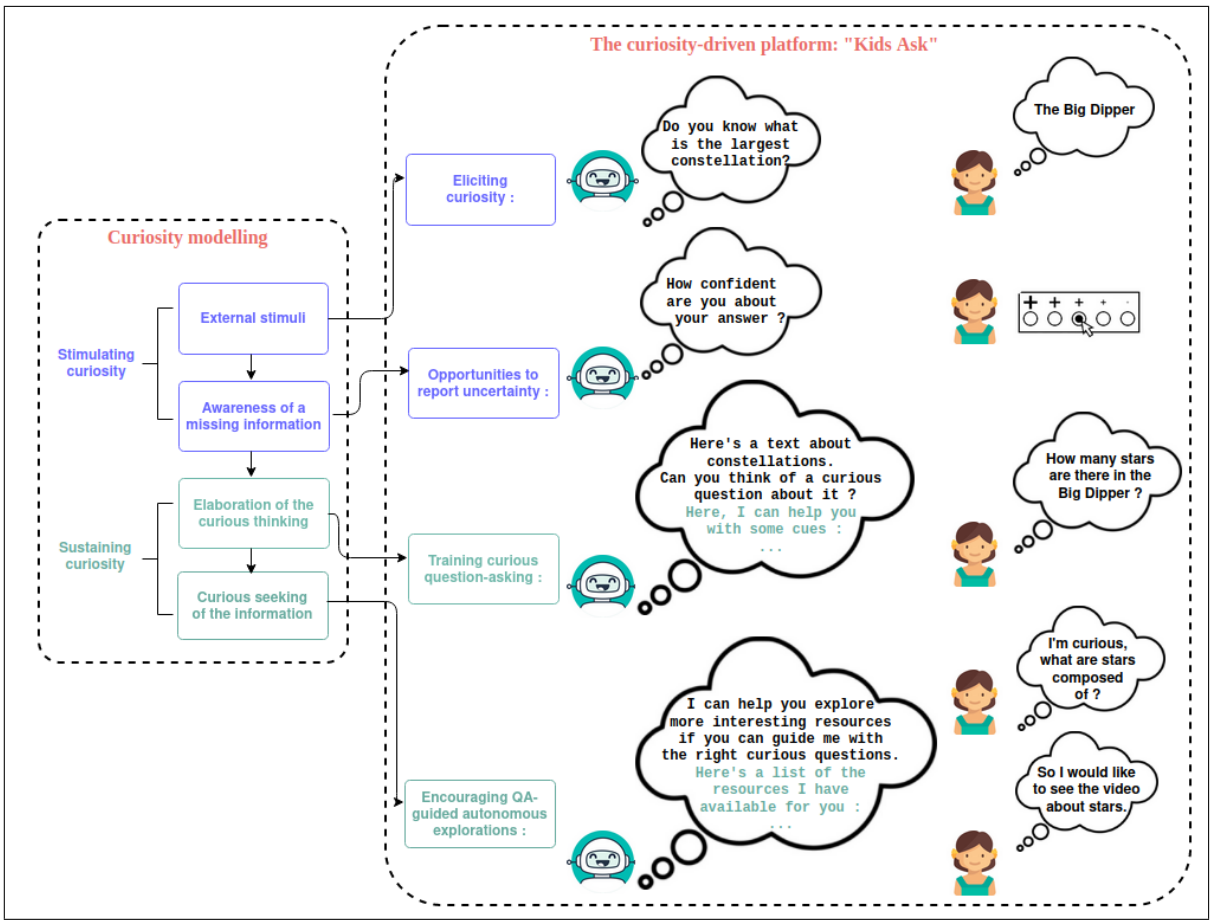


Figure 1: **Illustration of the study's general idea and the platform's main features**

2 Related work

2.1 Epistemic curiosity, meta-cognition and information-seeking behaviors

The main event that triggers individuals' information-seeking behaviors is them realizing that they are missing this information. Loewenstein explains this idea in his "information gap" theory [29] where he says that the activation of epistemic curiosity depends directly on the individual's awareness of his learning processes and how close he feels to closing a knowledge gap: "seeking out new knowledge is more motivated by the aversiveness of not possessing the information than it is by the anticipation of pleasure from obtaining it". Oudeyer and Kaplan, in [33], propose to complete this cycle with their Learning Progress Theory, suggesting that, to continue seeking information, individuals need to work on activities that provide them with a feeling of an optimal learning progress. Ultimately, Murayama [23] presents curiosity as an intrinsically-motivated, goal-oriented cycle that results in information-searching

behaviors and that is sustained by a continuous need to optimize the reward felt when acquiring new information. In [13], authors argue that it is also possible to feel a strong motivational urge for information, even in the absence of the expected pleasant experience that will emerge from acquiring it. They show that, together, the information's inherent reward and the cognitive desire for it can predict whether or not the information-seeking behaviors will be happening. Collectively, these theories suggest that training children's curiosity-driven learning should focus on training the cognitive mechanisms involved in its initiation and sustenance, such as the awareness of a knowledge-gap and how to use the appropriate strategies to resolve one.

2.2 Divergent questions and exploration during learning

In [16], authors investigate what motivates people to generate questions. Most importantly to our study, they highlight an epistemic curiosity-related factor: a deficit in knowledge and a need to compensate for it. Besides pinpointing the importance of question-asking in curiosity-driven processes, this finding also suggests the need to differentiate between question types: curiosity is more associated with questions that lead to learning new information than with those that are only used to scan explicit facts.

For this, Bereither [37] proposed two types of questions children can generate in a reading-comprehension task: text-based and knowledge-based. The first type are questions which answers can be directly derived from the text, while the second type reflects the child's cognitive desire to know or make sense of something. Although text-based questions are easier to think of, knowledge-based questions remain more interesting for learning as they involve higher-order thinking and more cognitive processes like the integration of newly-acquired information into one's knowledge base. Similarly, Gallagher and al. [14] propose to classify questions as divergent- and convergent-thinking. The first being surface-level questions that require children to explain or compare ideas while the second involve divergent thinking processes and require prediction, making hypotheses or judgements. Authors in [1] investigated the effect of curiosity on students' questioning abilities and found that the more curious a student is by trait, the more divergent-thinking questions he will ask.

On another note, in studying the role of question-asking during exploration and learning, researchers in [39] propose that asking a question and pursuing its answer is a fuel for inquiry-based and independent learning. Indeed, according to [25], finding the answer to a question requires the individual not only look for the information, but also to make inferences and use already-acquired knowledge in new situations. For example, in [6], authors developed a routine where participants learn by asking and exploring: they were first asked to formulate a 'critical' question related to the learning material. They were then required to explore the available resources and formulate subsequent questions. Finally, they were asked to use these questions and explored resources to formulate a concise answer to their 'critical' question. Results showed significant differences in learning progress between the students that were trained with this routine and those who had a traditional lecture session. These findings suggest that the explicit thinking of a question can positively affect students' answer-searching behaviors and learning performances.

Despite the important role of generating high-order questions during learning, this behavior is still almost absent from today's classrooms. In [16], the authors explain this by children's tendency to overestimate their knowledge: children don't ask questions because they don't know that they don't know. On another hand, Humphries and al. suggest in [20] that children don't ask higher-order questions because they don't know how to formulate them. In fact, their results showed that children tended to ask text-based and convergent-thinking questions because they are easier to construct. However, the results also suggest that the ability of asking divergent-thinking questions can be mentored through demonstrations and invitations to deeper thinking. Another brake keeping children from asking questions [34] is their negative perception of asking questions and the fear of classmates' judgement.

2.3 Technology applications for priming curiosity-driven behaviors during learning

Research has shown that the social environments have a paramount influence on children's exploratory behaviors and motivation to learn [36]. In this light, several studies have investigated the possibility to use social artificial agents to prime curiosity-driven behaviors in children. In Ceha et al. [7], children interacted with a social robot that either expressed curiosity, curiosity plus rationale, or no

curiosity during a rock classification game. Their results showed that children were able to recognize the robot's curiosity and to 'catch' its curiosity-driven emotions and behaviors. But in terms of intrinsic motivation, authors were not able to assess a significant difference between the conditions with respect to enjoyment, competence, effort and relatedness. In Gordon's work [15], children interacted with an autonomous agent that exhibited curiosity-driven behaviors. Results showed that the interaction selectively increased children's aspects of curiosity behavior, but led to no significant differences in learning progress, leaving the question of 'whether curiosity can lead to learning gains' still unanswered. In another study [38], Gordon also presents a question-asking based exploration game. The results showed that the precision of the questions that children generated was a significant predictor for their curiosity, as reported by their teachers. In addition, in Sher and al.'s work in [3], participants used a web-based application that favoured autonomous exploratory behaviors. Their results showed that enabling time to explore resulted in longer exploration and better learning of the self-explored knowledge, suggesting that exploration improves independent learning. On another note, Clément and al. [10] used the learning progress theory of curiosity to adapt learning tasks to each child and showed that students were more motivated to learn when the lessons were scheduled by curiosity-driven algorithms rather than with the predefined learning sequences. Finally, Alaimi and al. [1] developed a pedagogical agent to train children to generate either convergent or divergent questions, using specific semantic cues. These cues consisted of giving the answers to either divergent or convergent questions and asking participants to formulate the questions that lead to them. Their results showed that the divergent-thinking oriented support helped children increase their number of higher-level questions. In addition, and although they did not perceive a significant effect of the participants' perception of curiosity, they found a positive effect of this latter on the children's ability to generate divergent questions. As promising as their results are, the agent here does not move forward with its training to show children how to use this question-asking skill during their learning. They also do not inform us about the impact of this training on the children's domain-knowledge learning achievements.

2.4 Current study

Previous work motivates our study as it shows that technology-mediated solutions for enhancing curiosity-driven learning are a promising path. However, the question remains unanswered as to how could children use curiosity-based behaviors as a strategy to optimize their learning. Thereby, our study aims to design an attractive technology-mediated curiosity training and analyze its effect on children's learning strategies and progress. With this mindset, we implement a conversational agent that trains divergent-thinking questioning through fostering knowledge-gap awareness. It also encourages using this skill to lead intrinsically-motivated explorations that help gain new information. The agent finally proposes evaluations to help investigate the effect of the training on the domain-knowledge learning progress.

3 Study design

3.1 Kids Ask

"Kids Ask" is a novel web-based educational platform that involves an interaction between a child and a conversational agent. The platform is designed to teach children how to generate curiosity-based questions and use them in their learning in order to acquire new knowledge.

To meet this aim, "kids Ask" offers the three following workspaces:

- **The question-asking training space (The QA training space)**: In this space, the child has different texts relating to a theme of his choice. For each text, he interacts with the agent that will try to help him think of divergent questions by giving him specific cues. These cues are either a questioning word (example: 'what difference'), or a questioning word combined with an answer to a possible curiosity-driven question about the text (example: 'what difference' + 'The vaccine prevents the disease, the medication treats it'). The child is then asked by the agent to use the cue(s) to find the corresponding question (for the example above, the question could be 'What is the difference between vaccines and medication?'). The agent chooses to give one or both cues depending on the condition the child is assigned to: control or experimental (see section 3.2 for more details).

- **The exploration space**: In this space, the child has a library of animated educational videos; only 3 of these are initially accessible. Every time the child watches a video, the agent reappears and asks him to generate a divergent question about the video he just seen, without giving any cue. The more divergent questions the child generates, the more new videos he can unblock and therefore discover.

- **The evaluation space**: This space contains different domain-knowledge quizzes. The items can be skipped if the child decides that he doesn't know the answer. If, however, the child submits an answer, the agent asks him to report his confidence level in this answer with a 5-Likert scale (from "Super not confident" to "Super confident").

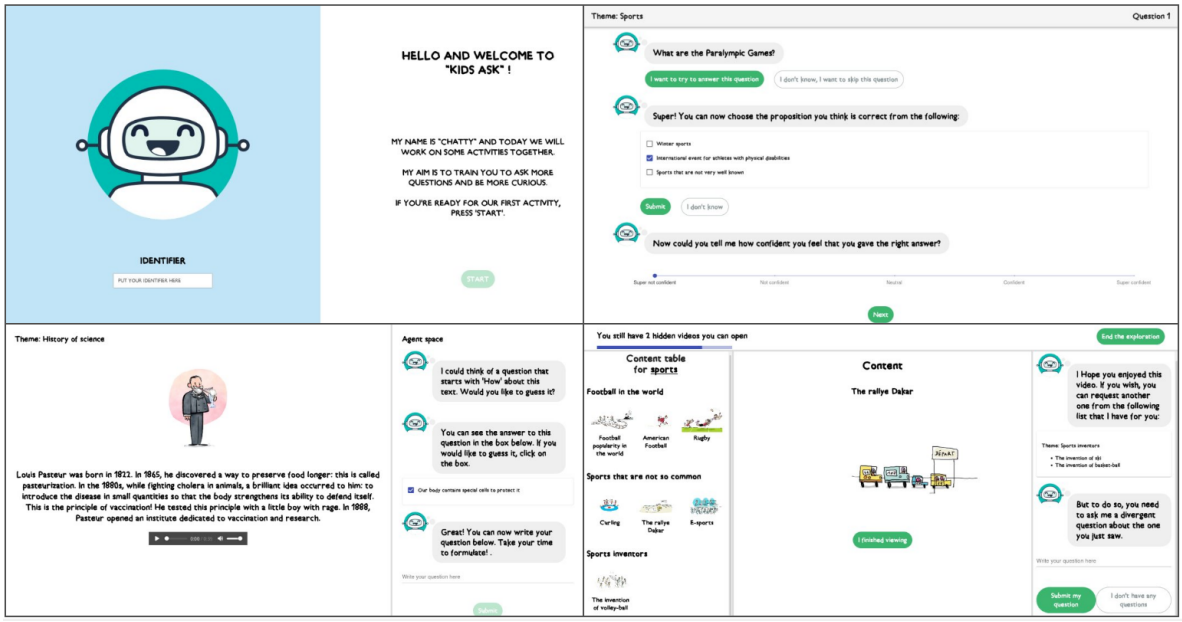


Figure 2: **Workspaces available in "Kids Ask": Homepage (Top left), The question-asking training wrokspace for the experimental condition (Bottom left), The evaluation workspace (Top right) and the exploration workspace (Bottom right)**

See Figure2 for the interfaces of the different workspaces and Section4.2 for bigger snips.

We organized the interaction in the same order for all participants: they all started with a general quiz relating to 6 themes then chose their favourite 2. They then went for the QA training space, followed by the exploration space. They also had a domain-knowledge evaluation quiz relating to the theme they chose for exploration both before and after this latter.

3.2 Experimental conditions

We had two conditions for this study. The difference between the two conditions resides only in the agent's behavior in the QA training space: For the control group, the agent helps the child only by giving questioning words that could start curiosity-driven questions about the text (example: 'What is'). Whilst for the experimental group, the agent also gives an answer to a possible curiosity-driven question relevant to the text and asks the child to think of a question that leads to this answer (example: the cues 'What is' and 'It is an illness caused by an intestine infection' are meant to lead the child to think of the question 'What is the Cholera disease?'.) In both cases, the child was free to either choose to use these propositions or to think of his own question. See table 1 for explicit examples of the difference in the agent's behavior.

In the exploration and evaluation spaces, the agent exhibited the exact same behavior for both conditions.

Control condition	**Agent**: "I could think of a question that starts with **'What difference'** about this text, try to guess it!. **Agent**: "Remember, the answer to my question is not in the text. Take your time to formulate!"
Experimental condition	**Agent**: "I could think of a question that starts with **'What difference'** about this text. You can see the answer to it in the box below. If you would like to guess it, click on the box." **Agent:** Renders the checkbox **'The vaccine prevents the disease, the medication treats it'**. **Agent**: "Super, you can now write your question. Take your time to formulate!"

Table 1: **Illustration of the difference in the agent's dialogue between the two conditions**

3.3 Design rationale

For the QA-training space, We chose to focus our training on divergent-thinking questions rather than convergent-thinking ones as the former relate more to epistemic curiosity and involve higher-order thinking and more cognitive processes (as mentioned earlier in section 2.2). In addition, we choose this design based on Alaimi's work [1], which had both divergent- and convergent-thinking oriented semantic cues and showed that the agent was actually able to help children generate significantly more divergent questions by using the divergent-thinking oriented semantic cues. Finally, we choose to maintain the linguistic help for both conditions (the questioning word) to help overcome the linguistic difficulties children can face when doing a question-generation task (i.e. not being familiar with the syntactic construction of a question or how to formulate an interrogation), as shown in [16].

Regarding the exploration, the task was designed to investigate whether participants were able to maintain the divergent QA skill that was trained in the previous space (i.e. agents' ability to transfer this skill). We also wanted to investigate if the question-asking skill helped them initiate and maintain autonomous explorations. Furthermore, the data collected in this space also serves us as a behavioral measure of epistemic curiosity. In fact, we use the idea developed in [17] where the authors measure curiosity as the individual's willingness to 'work' for information. Since we constraint children to think of divergent questions if they want to reveal new educational videos, these questions become the effort participants make when the only reward they get in return is the information (i.e. the video) itself.

Finally, we use the same quiz before and after the exploration phase in order to assess the participants' learning progress and the impact of the curiosity-driven behaviors on it. The platform's quizzes are skippable and ask children to self-report their confidence levels in their answers. Indeed, research such as in [16] shows that children do not engage in information-seeking behaviors mainly because they fail to identify their uncertainties. We therefore use the quiz and the confidence-level rating to try to elicit their curiosity by raising their awareness about information they might be missing. In addition, work in [35] suggests that giving children the possibility to skip a question allows them to think deeper about what they do and do not know and to report low certainty.

3.4 Design Approach

We adopted a participatory design approach and included the different stakeholders in our process (i.e., the teachers and the students). To do so, and after developing a first prototype of "Kids Ask", we proceeded to validate it with three primary school teachers. They helped evaluate and propose

modifications to the pedagogical approach, the content of the application and the study course (number and length of sessions).

We then began testing this version with 8 students. We saw difficulties in reading the texts and understanding the task. We then proceeded to change the content to be simpler and added audios for every text in the QA training space to help overcome reading problems. Finally, we tested the new version with 8 other students and saw that they were able to do the task successfully.

3.5 Technical implementation

The interface was programmed in JavaScript using the REACT library and was connected to a RESTful API to publish and retrieve the interaction data. As presented in the diagram below (Figure3), the behavior of the agent in terms of selection of the adequate cue(s) to offer was predefined and hand-scripted: it was connected to a database containing the different text resources and every text had a sequence of linguistic and semantic cues linked to it. Depending on the child's condition, the agent's automaton composes the dialogue utterances in order to include the appropriate support. We changed the utterances between the questions to avoid repetition in the agent's dialogue: a replica is not executed if it has been delivered during the previous question [22].

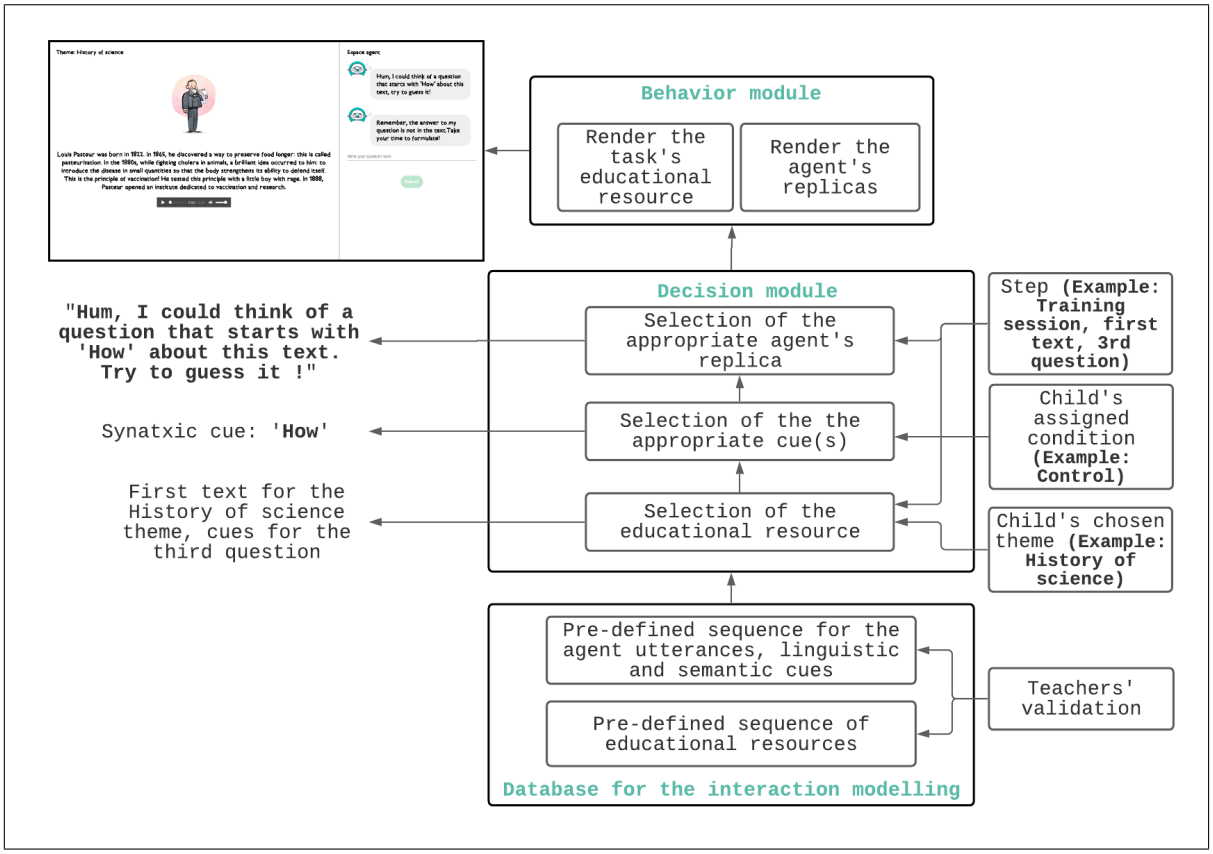


Figure 3: **System design**

4 Methodology

4.1 Participants

We recruited 57 CM1 students belonging to four classes from two French primary schools, they were between 9 and 10.5 years old. We were constrained to remove the data for 6 participants who had missing or unusable data (entered incomplete or incomprehensible phrases). This left us with 51

Measure	Control group	Experimental group	p-values
Age	9.36 ± 0.43	9.31 ± 0.43	0.66
Device use frequency	30.3 ± 6.28	27.24 ± 7.59	0.12
Curiosity trait	29.3 ± 4.3	27.12 ± 4.8	0.1
Perception of curiosity	38.8 ± 8	37.5 ± 6.58	0.53
Reading ability	280.03 ± 50.5	272.94 ± 96.6	0.75
QA fluency	9.61 ± 7.38	8.24 ± 5.56	0.46

Table 2: **Profile measures for the experimental and control conditions**

participants that were assigned either to the control group (26 with 12 boys and 14 girls) or the experimental group (25 with 13 boys and 12 girls).

The groups were assigned with a pseudo-randomized method after collecting profile data regarding the age, the device use frequency, the curiosity-related perceptions and the curiosity trait, the reading score and the initial question asking fluency performance (see 4.3 for more details of the measures). Thereby, and as shown in Table 2, we had two balanced groups that were not different in terms of initial profile measures.

4.2 Procedure

The experiment consisted of four sessions of 1h15 within the same week: One for collecting the profile measures mentioned above, one for the divergent-QA training, one for the exploration and one for the post-intervention assessment. See study timeline in Figure 4 and details of the different measures in the data collection paragraph 4.3.

Our study was approved by the institute's ethics committee (certificate n°2019-23) and only started after having all participants parents' signed consents.

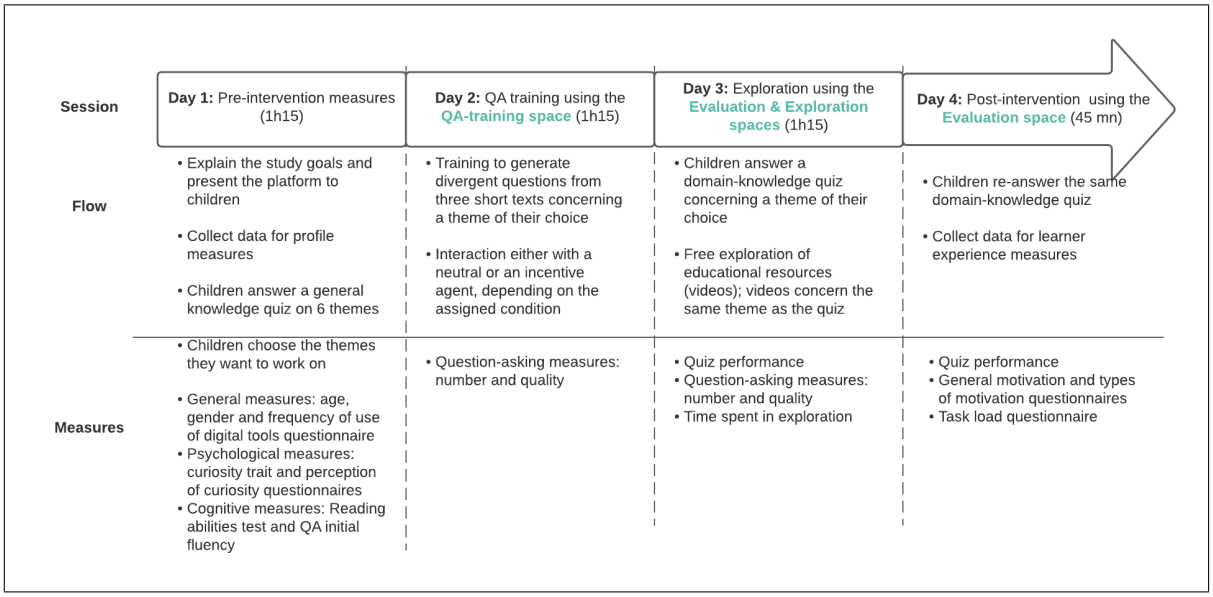


Figure 4: **Study timeline and measures**

Session 1 During the first session, we presented the study to the participants and collected the age and gender data. We also administered the questionnaires concerning the curiosity trait, the perception of curiosity and device use frequency. We then run the reading abilities test and the question-asking fluency test (see more details of the tests in the chapter 4.3). We also took time to explain what divergent questioning means and highlighted the difference between divergent and convergent questions. Participants then went to the evaluation space of "Kids Ask" and began the interaction with a general quiz (pre-test) relating to six themes. For each question, children had the

choice to either skip the item, by clicking on the 'I don't know, I want to skip this question' button, or answer it. If they do choose to answer, they are asked to report their confidence level in their answer: they had a 5-Likert scale from 'Super not confident' to 'Super confident' (see Figure 5).

Once this pre-test completed, children got to choose their two favourite topics and were told that they will work on these latter during the upcoming sessions.

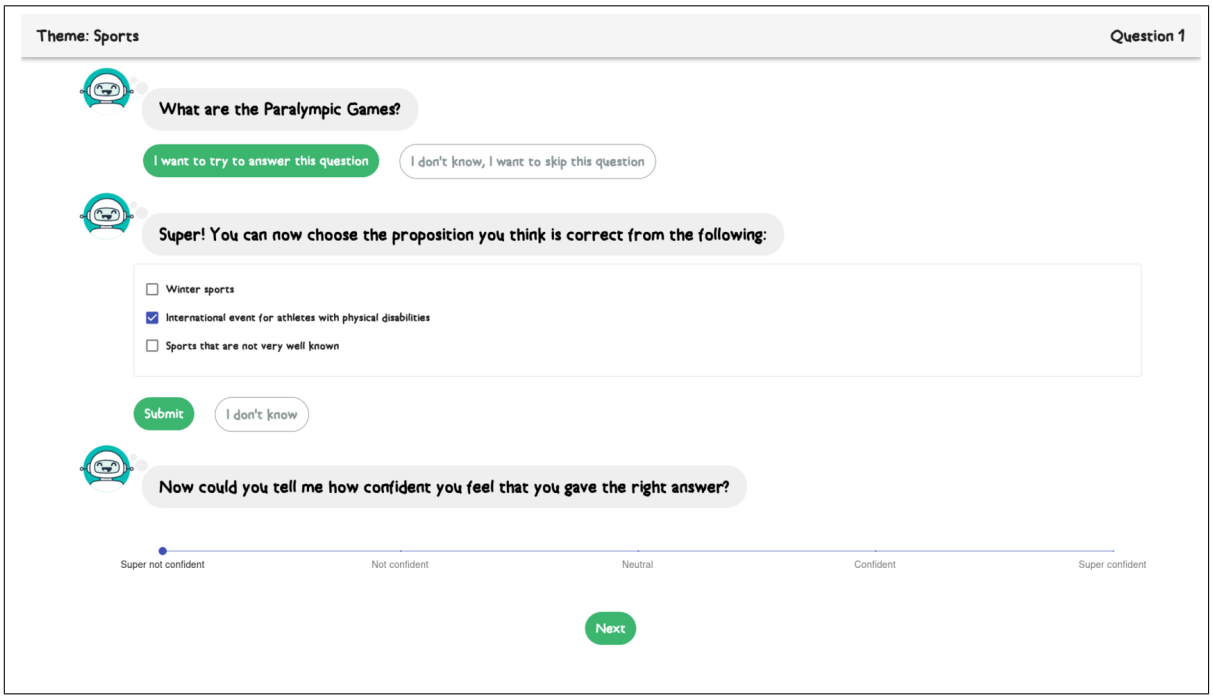


Figure 5: **Interface for the evaluation workspace**

Session 2 During the QA training session, participants were asked to read short texts relating to one theme they chose. They then started working on generating divergent-thinking questions relating to these texts using the cues provided by the agent. They also had audio players for each text to help them if they had reading difficulties. The texts were selected from online resources and children magazines (Sciences et vie Junior, Quelle Histoire and Questions? Réponses!) and were edited in order for them all to have six sentences and the same number of words. The average number of words per text was of 109.

To begin working on a text, participants were asked to read or listen to it and then click on the 'I finished reading' button once they understand it. This button enables the 'discussion' with the agent, in the agent's space on the right section of the screen. As explained above, the agent helps the child generate divergent-thinking questions about the text by giving either linguistic support (control group) or linguistic plus divergent-thinking semantic support (experimental group). See Figures 6 and 7 for the difference between the conditions.

During this session, children had to process three texts and generate six questions per text with the agent, making a total of 18 questions. They were not restricted in terms of time, apart from the session length. They had the application running on tablets and worked individually.

Session 3 Participants started with a quiz concerning the second theme they chose and then moved on to the exploration phase. During the exploration, they were encouraged to navigate autonomously amongst several educational videos but were required to use their divergent-questioning skills to do so. The videos were all taken from the same french website 1 jour, 1 question. They were all between 1mn20 and 1mn30 of length and contained the same amount of information and the same number of sentences and drawings.

At the beginning of exploration, participants had 3 educational videos related to their second theme in their 'Content table' section, and a progression bar suggesting that there are 6 more hidden ones. When a participant selects a video from these three, it appears in the central 'Content' section and he is able to play it. He is then asked to click on the 'I finished viewing' button once he finishes playing it.

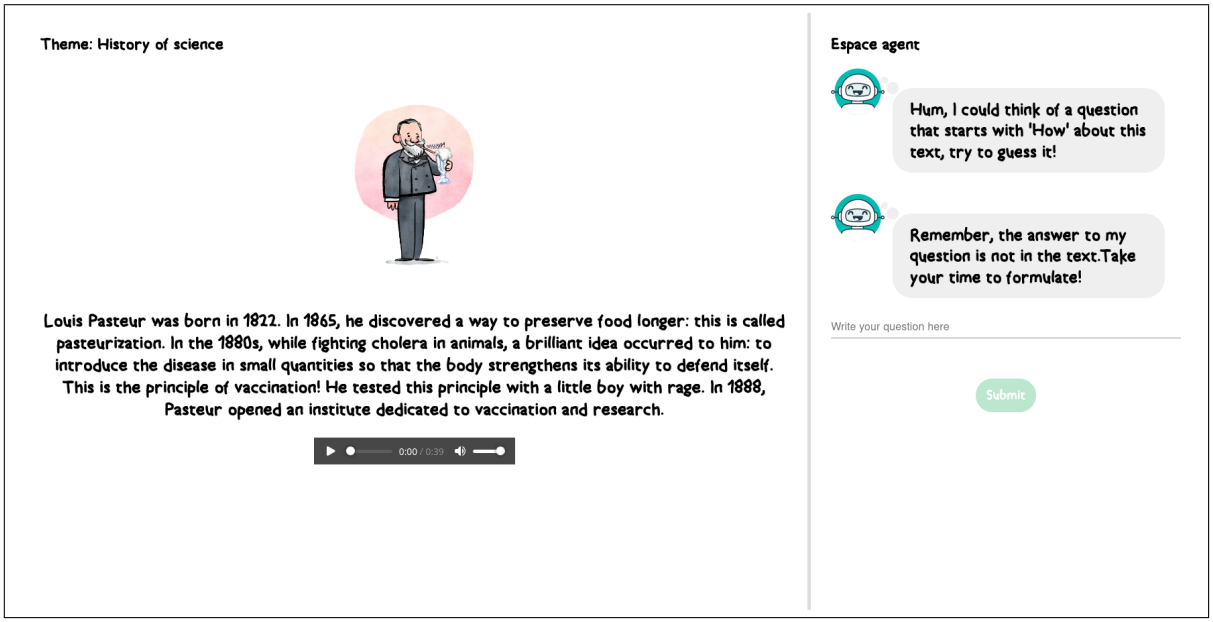


Figure 6: **Agent behavior for the control condition during the QA-training session**

This button makes the agent reappear in the right section of the page. It now shows the participant a list of the remaining hidden videos he can unblock for him (see Figure 8). It then informs him that in order to choose a new video to add to his content table, he is required to enter a divergent question about the one he just saw. If the child decides that he has no questions, the agent has no actions and the space remains unchanged. If, on the other hand, the child submits a question, he can choose a new video to open from the list. Once chosen, the video is added to his pedagogical content and becomes available for watching. The progress bar continuously indicates the remaining number of locked videos (See Figures 9 and 10 for the agent's behavior before and after the participant enters a question and chooses a new video).

There is a total of three initially-unlocked videos (children could navigate between them without generating questions), six new videos to unlock and 9 questions to generate over this session. Children were able to deliberately stop this session by clicking on the 'I finished exploring' button that appears on the top right of the screen, after opening at least three videos.

Session 4 Participants re-answered the same quiz they had during the pre-exploration. They also answered the post-intervention surveys for the general motivation, types of motivation and the task load (see measures timeline in 4 and further descriptions in the "Data collection and instruments" chapter 4.3).

4.3 Data collection and measures

4.3.1 Profile Measures

This includes general measures: age, gender and the technology frequency use, psychological measures: curiosity trait ([28]) and perception of curiosity using the CIAC questionnaire ([34]). And finally, two cognitive measures: the reading abilities using the standardized test in [27] and the question-asking fluency: this is an offline test where we give participants a short text to read then ask them to write as many questions about it as they can, within 2 minutes. The QA fluency measure consists of computing the sum of the scores of the questions generated during these 2 minutes, using the quotation grid that assesses the questions quality, described in Section4.3.2.

4.3.2 Behavioral curiosity measures

- *Percentage of divergent questions asked.* This measure consists of counting the total number of the correct questions generated during each session and computing the percentage of the curiosity-driven, i.e. the divergent, questions amongst them. We did not consider the correspondence

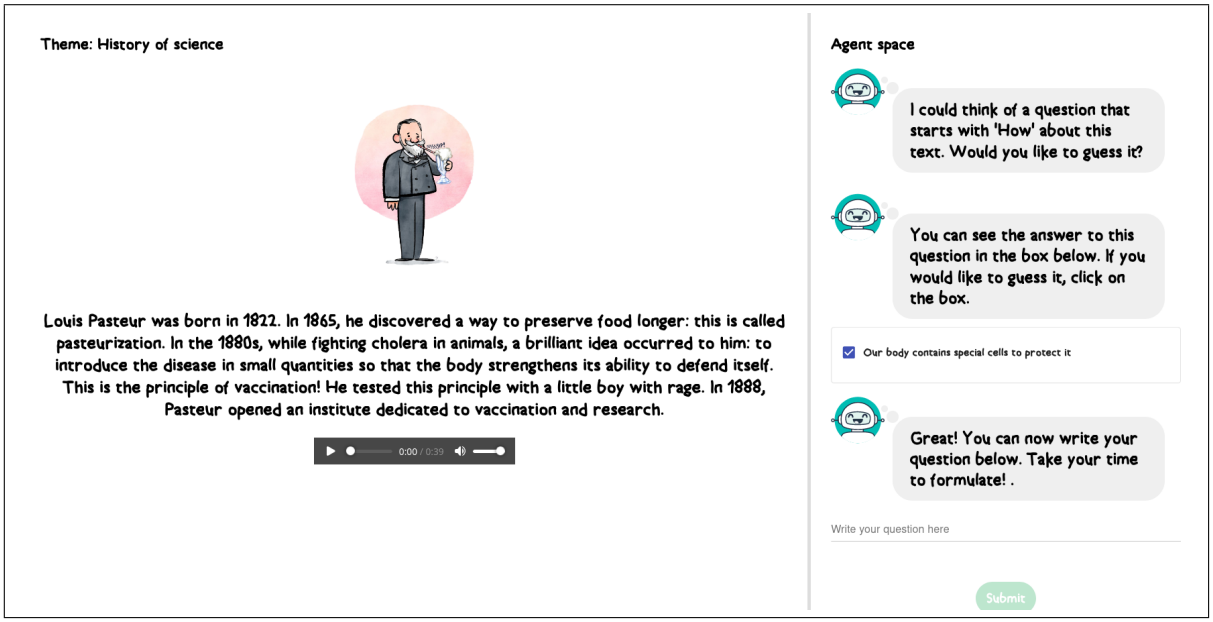


Figure 7: **Agent behavior for the experimental condition during the QA-training session**

between the question entered and the cues given by the agent, so children were free to imagine another question if they wanted to, as long as it is still related to the resource. If a question is repeated, it is only taken into account once.

A question is considered correct and taken into account if:

- It is a question — and not a statement.
- It is related to the educational resource (text or video).
- It is not repeated more than one time.

Example: For the text about Louis Pasteur (available in Figures 6 or 7), a linguistic cue '**What are the other**' and a semantic cue "**Fermentation**", we accepted questions such as:

- "What are the other possible ways to preserve food?" : used the agent propositions and is related to the text.
- "What are the other ways to protect our body?" : did not use the agent semantic proposition but is still related to the text.
- "How is the vaccine different from medication?" : did not use both of the agent propositions but is still related to the text.

However, data such as the following was not accepted:

- "It is another way to preserve food" : this is a statement and not a question.
- "What are the other parts of a robot?" : this is not related to the Louis Pasteur text.
- "What are the other characters of the Simpsons?" : this is not a serious attempt.

.

On another hand, a question was considered to be divergent if its answer is not explicitly stated in the educational resource (text or video) in question.

Example: The question "What are the other possible ways to preserve food?" is considered to be divergent, whereas "What are the other elements of the Pasteurisation process?" is considered convergent as the answer to it is explicitly stated in the text.

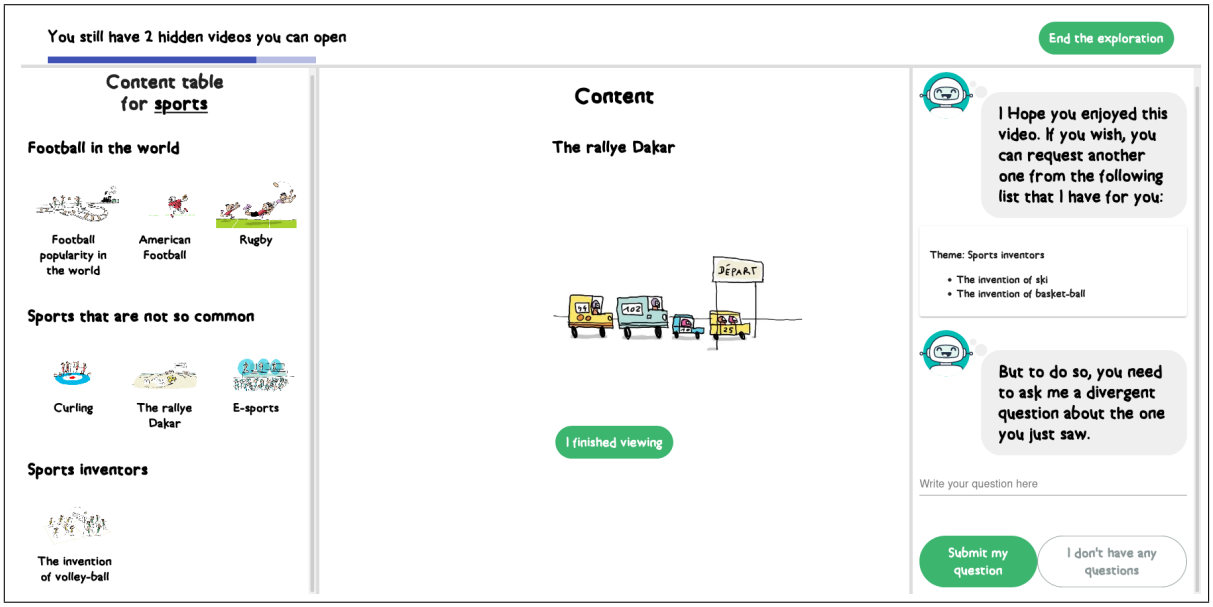


Figure 8: **Agent behavior after finishing viewing a video during the exploration**

This annotation process was performed by the two researchers who led the experiments in schools. All data was anonymized: coders could only see the identifiers that children were given randomly at the beginning of the intervention. The two researchers annotated 65% of the data generated during the the different phases. The inter-rater reliability was of 88,1%, with an agreement percentage of 90%. The rest of the data was only coded by the first author.

- *Quality of the questions generated* In scoring the quality of the questions generated, we used the grid that emerged from the classification in [14, 9] and computed a question's score as the sum of the following criteria:
  - One point if the question is high-level: the answer to this question is not a simple fact (example: 'How big is a dinosaur?') but requires to explain a mechanism, a relationship etc (example: 'Why were dinosaurs so big?').
  - From 1 to 4 points, based on the syntactic construction of the question :
    - 1 point for a 'closed' or declarative question (example: "Dinosaurs were big?").
    - 2 points for questions with questioning words in the middle of the sentence (example: "The dinosaurs were how big?").
    - 3 points for a question without an interrogative formulation (example: "Why the dinosaurs are big ?").
    - 4 points for a questioning word in the beginning of the sentence that has interrogative syntax (example: "Why are dinosaurs big?").
  - From 1 to 3 points, based on the use of questioning words :
    - 1 point for a declarative question i.e., with no questioning word (example: "Dinosaurs were big?").
    - 2 points for questions with 'Is/Are' (example: "Are dinosaurs big?").
    - 3 points for the use of proper questioning words (example: "How big were the dinosaurs?").

  Ultimately, questions had scores from 2 to 8 points. For each participant, we calculated an average score for all the questions generated during each session.

  Again, the same two raters coded 65% of the questions generated following this grid and had an inter-rater reliability of 79.6%, with an agreement percentage of 84.8%. The rest was coded by the first rater.

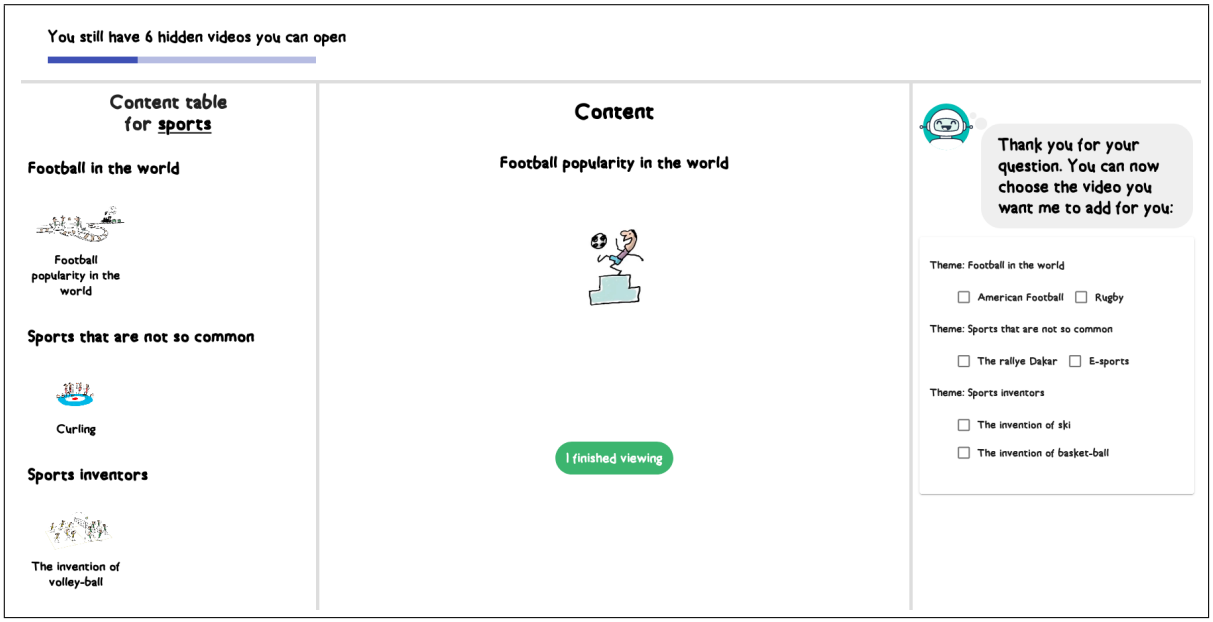


Figure 9: **Adding the new video "American football" to the content table**

- *Time spent in exploration.* This metric reflects the number of videos explored. We consider this as a curiosity measure as children were told that they are able to end the session if they feel that they are not curious about discovering the new resources anymore.

4.3.3 Domain-knowledge learning progress.

Participants had to answer the same domain-knowledge quiz before and after the exploration phase; the maximum score is of 9 points. The items were like the following:

- 1/3 of the questions had their explicit answers in the videos.
- 1/3 of the questions required the participant to link information from different videos to find an answer.
- 1/3 of the questions did not relate to the theme chosen for the exploration. These are our control questions to help us investigate if the learning progress made is actually related to the exploratory behavior.

The learning progress was then computed in order to evaluate the participant's progress with respect to the maximum learning gain achieved:

$$learning\ progress = \frac{Score_{postExploration} - Score_{preExploration}}{Score_{Max}} \tag{1}$$

With: $Score_{preExploration}$ and $Score_{postExploration}$ being, respectively, the scores during the pre- and post- exploration quizzes. And, $Score_{Max}$ being the highest score achieved in the post-exploration quiz.

4.3.4 Learner experience measures

- *Motivation measures.* The participants' motivation to use "Kids Ask" was assessed via two questionnaires: the general motivation scale [11] and Vallerand's types of motivation scale [40].

  The general motivation scale was used to investigate the potential short-term motivation. It contains one sub-scale for evaluating participants' satisfaction with the platform, one sub-scale for evaluating their perceived competence and one sub-scale for assessing the degree of preference with respect to a favourite school activity. The items were answered either with a 6- or 7-Likert scale, with the maximum score being 54.

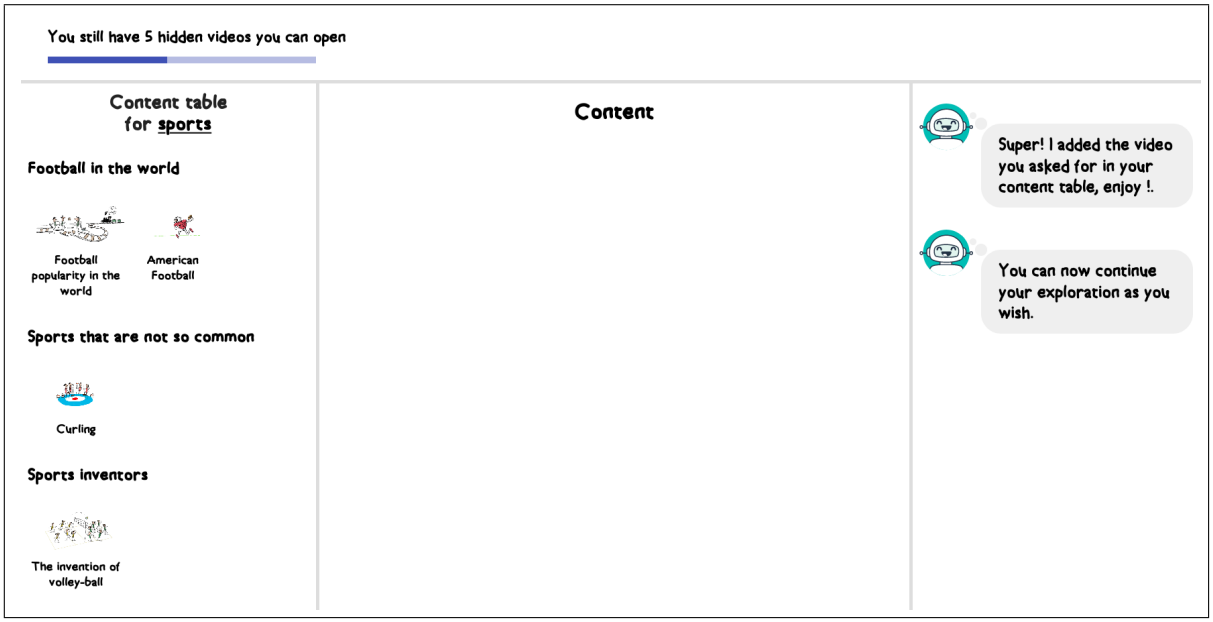


Figure 10: **Agent behavior after submitting a question during exploration**

On another hand, Vallerand's scale was used to probe intrinsic and extrinsic motivational mechanisms in the educational settings we had. It is composed of three sub-scales that differentiate: intrinsic motivation (possible scores from 0 to 9 points), extrinsic motivation (possible scores from 0 to 9 points) and amotivation (possible scores from 0 to 3 points). All questionnaire items are yes or no questions.

- *Task load.* The participants' behavior was also evaluated in terms of the subjective workload they experienced during the intervention. For this, we used the NASA-TLX workload multi-dimensional scale developed in [18]. Information about the intensity of six workload-related factors are used in order to estimate a reliable measure of workload: mental demand, physical demand, temporal demand, performance, effort and frustration.

4.4 Research questions and hypotheses

The main goal of this research is to design a conversational agent that can train children's divergent-thinking question-asking skills and foster the use of this skill to maintain autonomous exploratory behaviors that lead to domain-knowledge learning progress. We hypothesize that:

- The incentive agent, compared to the neutral agent, increases children's ability to generate divergent-thinking questions.
- The incentive agent, compared to the neutral agent, increases children's ability to maintain autonomous explorations.
- The incentive agent, compared to the neutral agent, leads to stronger domain-knowledge learning progress.
- Differences in the achieved domain-knowledge learning progress can be explained by differences in curiosity-driven behaviors.
- The incentive agent, compared to the neutral agent, will be associated with a better learning experience and a higher intrinsic motivation.

5 Results

To address our research questions, we conduct the data analysis so that we are able to distinguish the effect of interacting with the two agents on children's divergent question-asking abilities. We also

analyze if children were able to maintain this skill in the absence of the agent support and use it to lead autonomous exploratory behaviors. On a further step, we investigate the effect of these curiosity-driven behaviors on the domain-knowledge learning progress. And finally, we look into the learners' experience with the platform in terms of cognitive load and and motivation of use (Figure11).

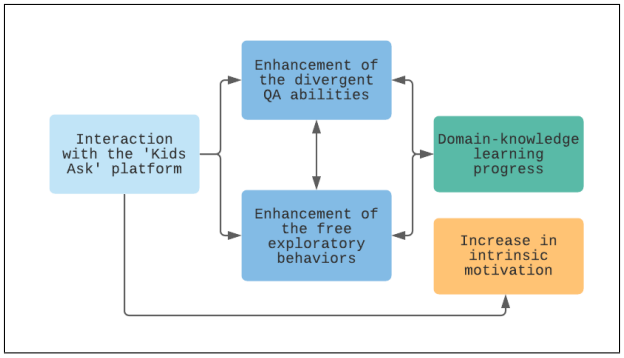


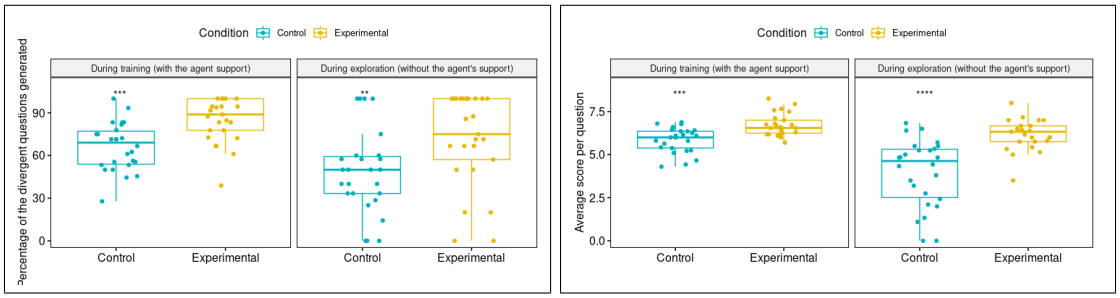
Figure 11: **Relationships to investigate with the "Kids Ask" platform**

5.1 Behavioral curiosity measures

5.1.1 Divergent question-asking performance and effect of the training

In analyzing the percentage of divergent questions generated, we used a two-way mixed ANOVA test, followed by Post-hoc tests with the Benferroni procedure. Our aim being to evaluate if the performance was different between the two conditions, between the two sessions and, finally, if it depended on the agent's behavior in the same way during both sessions. Results showed a statistically-significant effect of the condition nature on the percentage of divergent questions generated: $F(1,49)=17.87$; $p=0.0001$); see sub-figure (a) in Figure12. On another hand, the effect of session nature on the divergent QA percentage was also significant ($F(1,49)=10.81$; $p=0.002$). The post-hoc tests investigating the simple main effect of the session for each condition level revealed a significant interaction only for the control group: $F(1,49)=7.22$; $p=0.01$; the performance dropped drastically. This was not the case for the experimental group: $F(1,49)=3.61$; $p=0.07$ where the performance tended to stay stable. Finally, and as expected, the two-way interaction between the condition and the session on the performance was not significant ($F(1,49)=0.19$; $p=0.67$) suggesting that the divergent QA percentage depended on the agent's behavior in the same way during both sessions.

On a final note, in terms of quality of the questions generated, we used the grid described in section 4.3.2 and calculated an average score per participant and per session. Similar to the previous analysis,



(a) **Participants with the incentive agent generated significantly more divergent questions. Their performance did not drop significantly during the second session, revealing a stronger efficiency for this agent in transferring the divergent QA skill.**

(b) **Participants with the incentive agent generated significantly better divergent questions. Their performance did not drop significantly during the second session, revealing a stronger efficiency for this agent in transferring the divergent QA skill.**

Figure 12: Question-asking measures during the two phases for the two groups

results of the two-way mixed ANOVA showed that the assigned condition affected the average score of the questions: F(1,49)=37.48 and p<0.0001; see sub-figure (b) in Figure12. The output of our test also showed a significant two-way interaction between the condition and the session on the average score of the questions (F(1,49)=9.5; p<0.0001) indicating that the positive impact of the sessions on the questions quality depended on the condition nature.

Before running the tests above, assumptions about normality, absence of outliers and homogeneity of variance were verified.

Taken together, these results support the hypotheses that our incentive agent was more successful in training children's faculties to generate divergent questions both in terms of quantity and quality. It was also more successful in facilitating the transfer of this strategy and its use in new contexts, where no cues are given to help.

5.1.2 Exploratory behavior and effect of the training

This metric reflects the number of resources that children were interested in seeing during their exploration. As shown in Figure13, children who interacted with the incentive agent ended up spending more time exploring the educational resources than those who interacted with the neutral agent: m=1222.95 s, SD= 439.76 for the control group and m= 1593.15 s, SD= 634.81 for the experimental; with p-value = 0.02 and an effect size (Cohen's d) of 0.68.

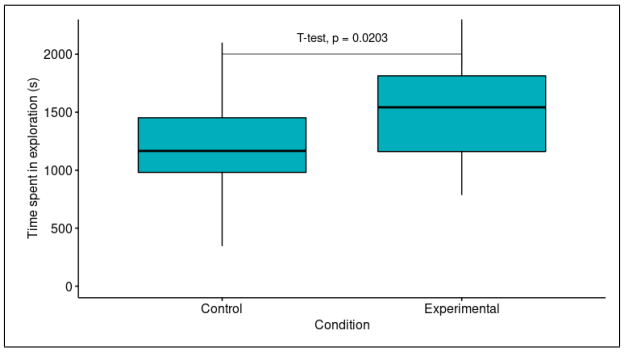


Figure 13: **Participants with the incentive agent spent significantly more time exploring the educational resources**

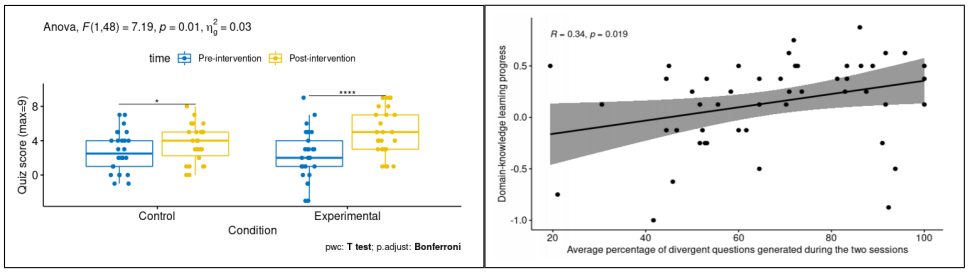
In analyzing the role of the training session on this difference, we run an ANCOVA test between the two conditions, with the time spent in exploration as a dependant variable and the divergent QA performance during the training as a co-variate. Results showed, indeed, a significant interaction: F(1,48)=4.1 and p=0.04. Before running the test, hypotheses about linearity between the response variable and the co-variate, normality and absence of outliers were verified.

Together with the results of the previous section, we can suggest that our incentive training was more successful in enhancing divergent QA faculties that, in their turn, had a significant positive effect on children's ability to lead autonomous explorations.

5.2 Domain-knowledge learning progress and effect of the divergent question-asking performance

In analyzing the two groups' learning progress, we compare participants' performances in the domain-knowledge quizzes before and after the exploration. It is to be noted that children's scores were initially similar for the two conditions, meaning that they both had the same space to achieve progress: m=2.73± 2.34 for the control group, m=2.32± 2.91 for the experimental group (maximum score =9) and p-value= 0.58.

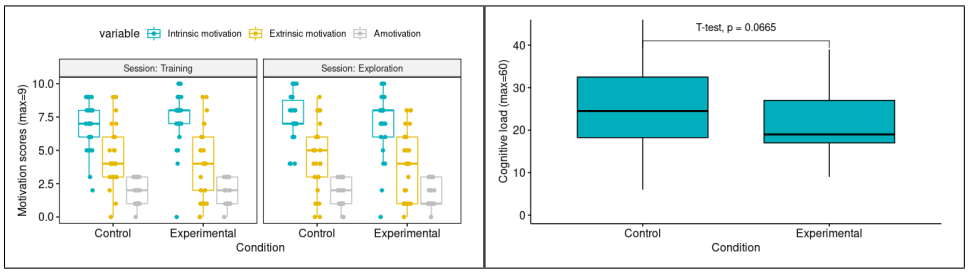
As shown in Figure14, the experimental group had a higher learning progress : m=0.25± 0.41 compared to m=-0.009± 0.45 for the control group; p-value= 0.04. Furthermore, we conducted a two-way repeated measures ANOVA test to see if the difference between the initial and final performances was significantly different between the two conditions. The interaction between time and the nature of the condition on the domain knowledge score came out statistically significant: F(1,48)=6.42 and



(a) **Participants with the incentive agent had a significantly higher learning progress even though they had similar initial scores**

(b) **The difference in learning progress is explained by the differences in the divergent QA performances**

Figure 14: Domain-knowledge progress and effect of the question-asking skills



(a) **Participants were significantly more intrinsically than extrinsically motivated during the two sessions but did not differ between the two groups**

(b) **The tasks load tended to be slightly lower for participants with the incentive agent**

Figure 15: Learner experience measures

p=0.015. Hypotheses concerning the absence of outliers and normality of the learning progress variable were verified pre running the test.

Finally, in analyzing the the role of the divergent QA performance on learning, we run an ANCOVA test between the conditions with the learning progress as a dependant variable and the average divergent QA performance for both sessions as a co-variate. The interaction was statistically significant with F(1,48)=7.19; p-value=0.01. The ANCOVA conditions were also verified.

With these results, we saw that the two groups were different regarding the learning progress that they achieved autonomously, and that this difference was mediated by how much divergent questions they were able to generate.

5.3 Learner experience measures

5.3.1 Motivation measures

The scores of general motivation did not differ between the two conditions (p=0.54) but remained high throughout the two sessions (m=39.3 for the experimental and m=37.74 for the control; max score =54). The difference between the groups in terms of intrinsic motivation was not significant either. However, and as shown in the sub-figure (a) in Figure15, participants were significantly more intrinsically than extrinsically motivated for both groups and during both sessions.

5.3.2 Tasks load

As seen in the second sub-figure in Figure15, the average tasks load scores for the two sessions tended to be slightly lower for the experimental group, suggesting that even with removing the agent's support, the task of generating divergent questions was perceived to be easier for participants of this group.

6 Discussion

In this work we investigate the efficiency of a conversational agent designed to foster curiosity-driven learning in children. For our data analysis, we rely on behavioral measures of epistemic curiosity such as the ability to generate divergent questions and to lead autonomous explorations.

Our findings showed that our incentive agent was more successful in enhancing divergent question-asking abilities and transferring them to new contexts where no help/ cues are given to help participants. We could therefore argue that children of the experimental group were more able to maintain their performance level not only because the agent was successful in training the formulation of divergent questions but, even further, because it was successful in training their abilities to identify knowledge gaps while learning. Indeed, this skill is considered as the fuel for an individual's curiosity-driven behaviors: not being able to identify missing information is considered as the first brake that keeps children from asking questions in classrooms [16].

In addition, our findings showed a better exploratory performance for the experimental group, with a significant impact of the divergent question-asking performance on the exploration time. The incentive agent's success in promoting longer question-guided explorations could be traced back to its behavior that relied on giving time and encouraging children to think of missing information, and write them down, while exploring. Indeed, this is similar to using a 'Thinking aloud' approach that capitalizes the idea that verbalizing one's thinking process explicitly while trying to solve a problem may lead to more carefully-planned problem-solving steps ([19], [2]). Thereby, and together with the QA-training session, we could suggest that our intervention succeeded in giving participants the essential metacognitive tools they needed to pursue a cycle of curiosity-driven learning: a training to get used to stopping and taking time to think of a potential knowledge gap (the exploration session), and a training to identify a knowledge gap and express it (the QA-training session).

These results are also in accordance with literature in meta-cognition that has repeatedly shown that learners often tend to end their learning cycles prematurely because they overestimate their mastery of the skills ([26], [31]). With our findings, we can suggest that training the faculty to stop, think of and ask divergent questions while learning might be a valuable strategy to be fostered in children in order for them to avoid the 'illusion of knowledge' trap and to prevent them from interrupting their information-seeking cycles prematurely.

As for the learning progress, we find that the experimental group was more successful and that our main behavioral curiosity measure (i.e. divergent question-asking performance) had a significant impact on it. This finding reinforces work such as in [32] that suggests a causal and bi-directional link between curiosity and learning. It is also in accordance with studies such as in [24], where authors show that states of curiosity can enhance learning and memory retention. Indeed, we can see that, because our incentive agent was more successful in eliciting participants curiosity, it helped them generate the appropriate divergent-thinking questions while exploring which, subsequently, led them to acquire new information. Finally, and as differences in the participants' learning progress were shown to be correlated to the differences in their divergent question-asking behaviors, we can suggest that monitoring these latter could be a valuable indicator for teachers to prognosis their students' learning progress.

Overall, and to the best of our knowledge, our study is the first that investigates the different components of a curiosity-driven learning loop: it brings an existence evidence to the links between knowledge-gap experiences, knowledge-seeking behaviors and learning progress. It also reinforces the support to the "information gap" theory.

As a final important observation, our results failed to show higher intrinsic motivation scores for the experimental condition. However, we noted stable and high general motivation scores, irrespective of the condition. This is counter our hypothesis that relies on the learning progress model where an increased learning progress should yield increased intrinsic motivations. Indeed, as our incentive agent generated more learning progress, we expected higher intrinsic motivation for its group. Several explanations can be advanced. First, and as mentioned in [16], the learning by asking questions strategy, especially using digital tools is still rarely used in classroom settings, even today. It is therefore possible that the activity's novel and playful characters alone have made it very motivating and attractive for children, with both agents. Second, the use of a participatory design approach could have had inter-played with this finding also. Such a design method is well documented to increase the match between the proposed solution and the learners' needs [21]. Indeed, this approach led us to gather pedagogical content that is both appropriate in terms of learning objectives and accessible in

terms of presentation format. Such features can enhance the solution's effectiveness and the learners' intrinsic motivation [8]. Finally, another possible explanation is children's incapacity to self-assess their learning progress, due to their metacognitive immaturity [30]. In fact, learning judgment failures would reduce the reinforcing power of learning progress on intrinsic motivation.

With rather encouraging results, our curiosity-based interactive system succeeded in promoting knowledge-gap awareness through the practice of higher-level question-asking; a skill that helped participants shape their self-sustained exploratory behaviors and gain new knowledge.

7 Limitations and future directions

A limitation to this study is the lack of interactivity in the dialogue between the agent and the child. One direction to a future study is to have a more sophisticated agent that will be able to give children feedback about the questions they submit. We also aim to implement an agent that is able to to generate divergent questions so it can engage in a turn-taking interaction with the child. Another possible limitation of this platform may be its very accessible design. The less-formal presentation of the resources (audios, videos) could have formed favourable conditions for our incentive agent to work. This could give us general directions to replicate our protocol with different material and a less-accessible design.

Finally, one of our aims is to raise children's awareness about the importance of the curiosity-driven behaviors on their learning. But with this experiment, and even though we were able to see an improvement in these behaviors and in learning progress, we do not know if the intervention actually helped participants see the link between the two. A future direction for this work will be to add a space in the platform where users can find pedagogical tutorials and activities that explain how to use their metacognitive skills and curiosity-driven behaviors to boost their learning.

8 Conclusion

In this work, we contribute to the promotion of curiosity-based approaches that foster individualized learning, while exploring the social advantages of using artificial conversational agents. We show that curiosity-driven behaviors can be practiced with the agent and that these behaviors are related to the learning progress children can achieve. Our work and results motivate the implementation of such approaches both in classroom settings and e-learning environments.

Acknowledgements

This work has been funded by the educational technologies start-up EvidenceB and the French National Association of Research and Technology (ANRT). The authors also thank the teachers who participated in this study and the research team members who helped conduct the experiments in classes.

References

[1] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. *Pedagogical Agents for Fostering Question-Asking Skills in Children*, page 1–13. Association for Computing Machinery, 2020.

[2] Maria Bannert and Christoph Mengelkamp. Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. does the verbalisation method affect learning? *Metacognition and Learning*, 3:39–58, 2008.

[3] Keren Ben-Tov Sher, Michal Levi-Keren, and Goren Gordon. Priming, enabling and assessment of curiosity. *Association for Educational Communications and Technology*, 2019.

[4] Daniel E. Berlyne. Curiosity and learning. *Motivation and Emotion*, 2:97–175, June 1978.

[5] JL. Bethier and J. Hubert. Le cerveau prédictif - magistère mémorisation. Video, January 2021.

[6] J. A. Bulgren, J. G. Marquis, B. K. Lenz, D. D. Deshler, and J. B. Schumaker. The effectiveness of a question-exploration routine for enhancing the content learning of secondary students. *Journal of Educational Psychology*, 103(3):578–593, 2011.

[7] Jessy Ceha, Joslin Goh, Corina McDonald, Dana Kulić, Edith Law, Nalin Chhibber, and Pierre-Yves Oudeyer. Expression of curiosity in social robots: Design, perception, and effects on behaviour. *CHI Conference on Human Factors in Computing Systems (CHI2019)*, page 1–12, 2019.

[8] Ho A.G. Chau P.W. The influence of multi-media related to creative learning. *In: Nazir S., Ahram T., Karwowski W. (eds) Advances in Human Factors in Training, Education, and Learning Sciences. AHFE 2020. Advances in Intelligent Systems and Computing*, 1211, 2020.

[9] M. M Chouinard. Children's questions: A mechanism for cognitive development: Iii. diary study of children's questions. *Monographs of the Society for Research in Child Development*, 72:45–57, 2004.

[10] Benjamin Clément. *Adaptive Personalization of Pedagogical Sequences using Machine Learning.* PhD thesis, University of Bordeaux, 2018.

[11] D. I. Cordova and M. R. Lepper. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, pages 715–730, 1996.

[12] Patricia Cranton. *Understanding and promoting transformative learning: A guide to Theory and Practice.* Stylus Publishing, 2016.

[13] Lily FitzGibbon, Johnny K. L. Lau, and Kou Murayama. The seductive lure of curiosity: Information as a motivationally salient reward. *PsyArxiv*, 2020.

[14] J. J. Gallagher and M. J. Aschner. A preliminary report on analyses of classroom interaction. *Merrill-Palmer Quarterly*, 9:183–194, 1963.

[15] Goren Gordon, Cynthia Breazeal, and Susan Engel. Can children catch curiosity from a social robot? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, page 91–98, New York, NY, USA, 2015. Association for Computing Machinery.

[16] Arthur C. Graesser and Natalie K Person. Question asking during tutoring. *American Educational Research Journal*, 31:104–137, 1994.

[17] M. Gross, C. M. Zedelius, and J. Schooler. Cultivating an understanding of curiosity as a seed for creativity. *Current Opinion in Behavioral Sciences*, 35:77–82, 2020.

[18] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. pages 139–183, 1988.

[19] Lisa M Henjens. The use of think-aloud strategies to solve word problems. Master's thesis, University of Nebraska - Lincoln, Lincoln, NE, 2007.

[20] Jean Humphries and Molly Ness. Beyond who, what, where, when, why, and how: Preparing students to generate questions in the age of common core standards. *Journal ofResearch in Childhood Education*, 29:551–561, 2015.

[21] Ole Sejer Iversen, Rachel Charlotte Smith, and Christian Dindler. Child as protagonist: Expanding the role of children in participatory design. In *Proceedings of the 2017 Conference on Interaction Design and Children*, IDC '17, page 27–37, New York, NY, USA, 2017. Association for Computing Machinery.

[22] Aidan Jones, Susan Bull, and Ginevra Castellano. "i know that now, i'm going to learn this next" promoting self-regulated learning with a robotic tutor. *International Journal of Social Robotics*, 10:439–454, 2018.

[23] Murayama K., , FitzGibbon L., and Sakaki M. Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*, 31:875–895, December 2019.

[24] M.J. Kang, M. Hsu, I.m. Krajbich, G. Loewenstein, S.M. McClure, J.T Wang, and C.F Camerer. The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol Sci.*, (20(8)):963–73, 2009.

[25] Walter Kintsch. Text comprehension, memory, and learning. *American Psychologist*, 49(4):294–303, 1994.

[26] Nate Kornell and Robert A. Bjork. Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Journal ofExperimental Psychology: Learning, Memory, and Cognition*, 42:914–924, 2016.

[27] Pierre Lefavrais. *Alouette-R : [test d'analyse de la vitesse en lecture a********** partir d'un texte]*. Paris : Centre de psychologie applique**********e, 2005.

[28] Jordan A. Litman and Tiffany L. Jimerson. The measurement of curiosity as a feeling of deprivation. *Journal of Personality Assessment*, 82, 2004.

[29] G. Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116:75–98, July 1994.

[30] van Loon Mariëtte and Janneke van de Pol. Judging own and peer performance when using feedback in elementary school. *Learning and Individual Differences*, Volume 74:101–754, 2019.

[31] K. Murayama, A. B. Blake, T. Kerr, and A. D. Castel. When enough is not enough: information overload and metacognitive decisions to stop studying information. *Memory*, 16:2:125–136, 2008.

[32] Pierre-Yves Oudeyer, Gottlieb Jacqueline, and Manuel Lopes. Intrinsic motivation, curiosity and learning: theory and applications in educational technologies. *Progress in brain research, Elsevier*, (229):257–284, 2016.

[33] Pierre-Yves Oudeyer, Frédéric Kaplan, and Verena V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, 2007.

[34] Tim Post and Juliette H. Walma van der Molen. Development and validation of a questionnaire to measure primary school children's images of and attitudes towards curiosity (the ciac questionnaire). *Motivation and Emotion*, 42,1:159–178, 2015.

[35] C.M Roebers, N. von der Linden, and P. Howie. Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology*, 25:109–134, 2007.

[36] Engel S. Children's need to know: Curiosity in schools. *Harvard Educational Review*, 81(4):625–645, 2011.

[37] Marlene Scardamalia and Carl Bereiter. Text-based and knowledge based questioning by children. *Cognition and Instruction*, 9:177–199, 1992.

[38] Noam Tor and Goren Gordon. Digital interactive quantitative curiosity assessment tool: Questions worlds. *International Journal of Information and Education Technology*, 10:614–621, 2020.

[39] Ronald D. Vale. The value of asking questions. *Molecular biology of the cell*, 24,:680–682, 2013.

[40] R. J. Vallerand, M. R. Blais, N. M. Brière, and L. G. Pelletier. Construction et validation de l'échelle de motivation en éducation (eme) [construction and validation of the motivation toward education scale]. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 21:323–349, 1989.

[41] Virginia E. Vitiello, Jamie J. Jirout, and Sharon K. Zumbrunn. Curiosity in schools. In Goren Gordon, editor, *The new science of curiosity*. Nova Science Publishers, Inc, 2018.