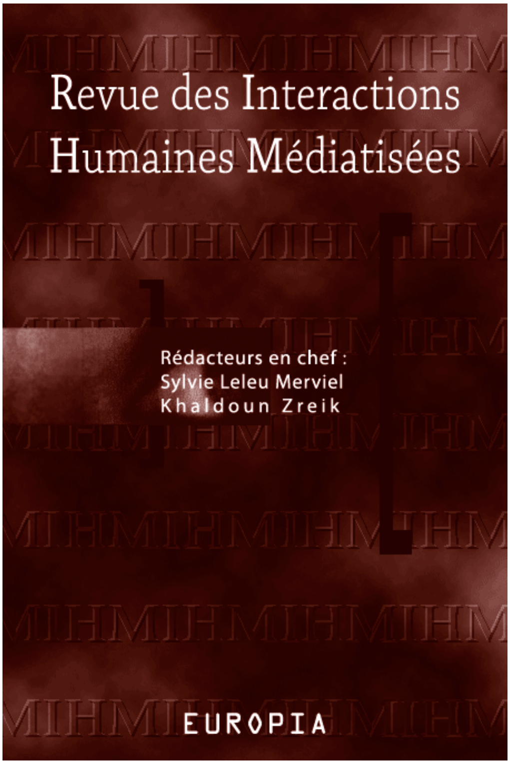
Revue des Interactions Humaines Médiatisées

Journal of Human Mediated Interactions

Rédacteurs en chef : Sylvie Leleu-Merviel & Khaldoun Zreik

Vol 22 - N°2/ 2021



© europia, 2021

15, avenue de Ségur,
75007 Paris - France

http://europia.org/RIHM | http://rihm.fr
Contact | e-mail : rihm@europia.org

Revue des Interactions Humaines Médiatisées

Journal of Human Mediated Interactions

Rédacteurs en chef / *Editors in chief*

- Sylvie Leleu-Merviel, Université de Valenciennes et du Hainaut-Cambrésis, Laboratoire DeVisu
- Khaldoun Zreik, Université Paris 8, Laboratoire Paragraphe

Comité éditorial / *Editorial Board*

- Thierry Baccino (Université Paris8, LUTIN - UMS-CNRS 2809, France)
- Karine Berthelot-Guiet (CELSA- Paris-Sorbonne GRIPIC, France)
- Pierre Boulanger (University of Alberta, Advanced Man-Machine Interface Laboratory, Canada)
- Jean-Jacques Boutaud (Université de Dijon, CIMEOS, France)
- Aline Chevalier (Université Paris Ouest Nanterre La Défense, CLLE-LTC, France)
- Yves Chevalier (Université de Bretagne Sud, CERSIC -ERELLIF, France)
- Didier Courbet (Université de la Méditerranée Aix-Marseille II, Mediasic, France)
- Viviane Couzinet (Université de Toulouse3, LERASS, France)
- Milad Doueihi (Université de Laval - Chaire de recherche en Cultures numériques, Canada)
- Pierre Fastrez (Université Catholique de Louvain, GReMS, Belgique)
- Pascal Francq (Université Catholique de Louvain, ISU, Belgique)
- Bertrand Gervais (UQAM, Centre de Recherche sur le texte et l'imaginaire, Canada)
- Patrizia Laudati (Université Côte d'Azur, SICLAB Méditerranée, France)
- Catherine Loneux (Université de Rennes, CERSIC -ERELLIF, France)
- Marion G. Müller (Jacobs University Bremen, PIAV, Allemagne)
- Marcel O'Gormann (Univerity of Waterloo, Critical Média Lab, Canada)
- Serge Proulx (UQAM, LabCMO, Canada)
- Jean-Marc Robert (Ecole Polytechnique de Montréal, Canada)
- Imad Saleh (Université Paris 8, CITU-Paragraphe, France)
- André Tricot (Université de Toulouse 2, CLLE - Lab. Travail & Cognition, France)
- Jean Vanderdonckt (Université Catholique de Louvain, LSM, Blgique)
- Alain Trognon (Université Nancy2, Laboratoire InterPsy, France)

Revue des Interactions Humaines Médiatisées

Journal of Human Mediated Interactions

Vol 22 - N°2 / 2021

Sommaire

Editorial

Ce numéro 22(2) de 2021 de R.I.H.M., *Revue des Interactions Humaines Médiatisées*, accuse, plus encore que les précédents, l'effet retard dû à la crise sanitaire. Les conséquences persistent pour notre revue : fort décalage avec la date calendaire de référence, et limitation à trois articles longs au lieu de quatre, conformément au nouveau régime instauré à partir du numéro 21(1). En outre, le numéro est atypique car les trois articles proviennent tous de la même unité de recherche, le laboratoire DeVisu (Design Visuel et Urbain) de l'Université polytechnique Hauts-de-France. C'est donc une sorte de revue de quelques travaux en cours dans cette unité de recherche. Cela n'exclut pas une grande variété d'objets, et surtout de méthodes.

En effet, le premier article est une étude idiographique visant à la quantification de données qualimétriques de type verbal. Pour cela, l'étude axiomatico-inductive se focalise sur un corpus constitué de 15 entretiens semi-ouverts portant sur le concept d'écosystème d'innovation. Le protocole explicite la transformation des enregistrements sonores des entretiens en matériau textuel via une analyse thématique. À partir de cette analyse, des matrices numériques sont élaborées par une logique dite triadique. Ainsi, l'étude met en évidence la pertinence, pour ceux engagés dans la recherche qualitative, d'une démarche qualimétrique fondée sur un processus d'auto-confrontation méta-réflexif.

A l'inverse, le deuxième article s'ancre sur une approche strictement statistique. L'expérimentation a été réalisée en 2017 en France et s'est basée sur les performances déclarées de 230 élèves de 28 classes de première année pour le codage, le décodage et la programmation. Cette étude vise à répondre à la problématique suivante : parmi les différentes modalités disponibles (corps, robot et tablette), quels types de performances ont été identifiés pour des enfants de cinq ans confrontés à la conception de programmes dans des classes de première année ? La méthodologie statistique employée permet de déterminer les modalités ayant les scores de performance les plus élevés.

Enfin, le troisième et dernier article est consacré à l'innovation pédagogique et mixe les approches qualitatives et quantitatives. L'objet de l'étude est de recueillir et d'analyser le retour d'expérience des 1899 étudiants concernés par le dispositif PRéLUDE pour la session 2021. Le recueil s'est fait par un questionnaire administré en ligne, occasionnant 350 réponses. Il comporte des données quantitatives, qui ont été analysées quantitativement, et des données qualitatives qui ont été analysées qualitativement. L'analyse des résultats montre que les répondants, perdus dans les informations du dispositif, ont du mal à donner un sens à ce nouveau dispositif qui leur paraît manquer de lien avec leur cursus et leur projet personnel et professionnel, malgré l'approche par compétences revendiquée par le dispositif et l'individualisation visée.

Nous vous souhaitons à toutes et à tous une très bonne lecture et nous vous remercions de votre fidélité.

Sylvie **LELEU-MERVIEL** et Khaldoun **ZREIK**
Rédacteurs en chef

An educational robotics experiment conducted with five-year old pupils to learn coding / decoding / design

Statistical results and interpretations

Julian ALVAREZ (1, 5), Katell BELLEGARDE (2, 5), Julie BOYAVAL (3), Vincent HUREZ (4), Jean-Jacques FLAHAUT (5), Thierry LAFOUGE (6)

(1) Laboratoire DeVisu, Hauts-de-France Polytechnic University, France
julian.alvarez@inspe-lille-hdf.fr
(2) Laboratoire CIREL, Lille University, France
katell.bellegarde@inspe-lille-hdf.fr
(3) Academy of Lille - District of Carvin, France
julie.boyaval@ac-lille.fr
(4) Academy of Dijon - District of Dijon Est, France
vincent.hurez@ac-dijon.fr
(5) INSPE (Academy of Lille), Hauts-de-France, Lille University, France
jean-jacques.flahaut@inspe-lille-hdf.fr
(6) Laboratoire ELICO, University of Lyon 1, France
thierry.lafouge@univ-lyon1.fr

Abstract. The main objective of this study is to assess the reliability of the Blue Bot experiment by statistical means. This experiment was carried out in 2017 in France and was based on the reported performances of 230 pupils in 28 Year-1[1] classes for coding, decoding and programming. 35 teachers were involved in the Blue Bot project. This study aims **at** reply**ing** the following **issue**: Out of the various available modalities (body, robot and tablet), what types of performances were identified when five-year-old children tackled decoding, coding and programming design activities in Year-1 classes? A statistical methodology was designed to ascertain which modalities had the highest performance scores. Th**at** allowed us to find **out** that all the groups using one or two combined modalities present**ed** a significant progression between the pre-tests and the post-tests. Th**o**se progressions were vastly superior to the Placebo (P) group. We also noticed that **the** preliminary activities and the various teaching modalities offered to the pupils during the experiment had a significant effect. This means that teaching sequential programming to Year-1 children is possible. At the level of comparative aspects related to the different modalities, in the case of the decoding and programming design activities, the Robot and Tablet modality (RT) presented the best performance rates. The Tablet (T) group presented the lowest progression rates in the experiment. After a brief description of the Blue Bot experience, this article sets out the statistical methodology, the results obtained and propounds an argument about it.

1 In this article, Y 1 means Year-1

Keywords. Robot-based pedagogy, modalities, digital tablet, comparative study, Year-1, experiment, coding, decoding, programming design, evaluation research, serious games.

1 Introduction

In France, the pupils targeted are children from three to five years of age. Computer programming is not included in the curriculum as we can find in other countries, for instance with the STEM (Science, Technology, Engineering, and Math) program which presents coding topics (Gribble et al. 2020). The only reference made of programming is the identification of the principle behind the algorithm and the pursuit of its application (*Bulletin Officiel* n°2, March 26, 2015). The **o**bjectives in the field of new technologies for pupils of five years old are rather geared towards "Using digital objects: cameras, digital tablets, computers". Thus, programming in kindergarten is presented as a means for children to work and acquire skills in other teaching areas, which are: oral communication, number sense, structuring time and space, problem solving, collaboration and abstraction (Greff, 2004).

Thus, to introduce computer programming to pupils, we had to search for ways to position ourselves with regard to French kindergarten programs. This is the reason why we opted for the "Explore the world" section, and more particularly the development of capacities related to spatial awareness and movement in space, which seemed to be the most suited to our research project.

Indeed, the French 2015 kindergarten programs suggested that these objectives could be met by using a variety of tools and by implementing a variety of teaching situations. The pupils should be offered four ways of learning according to the French National Education Ministry (French Official Bulletin (*Bulletin Officiel*) n°2, March 26, 2015):

- **Learning through games,** which allows pupils to have an impact on the real world by developing their imagination and autonomy, all of which encourages communication with each other.
- **Learning through thought and problem solving**: this was mentioned in the pedagogy of acting, succeeding and understanding.
- **Learning through practice**: pupils stabilize their knowledge when they practice; through practice, their understanding of the executed strategies leading to success are corroborated, thus allowing them to measure their progress.
- **Learning through remembering and memorizing**: providing regular feedback on what pupils already know allows them to strengthen their skills and to keep improving.

By applying these four ways of learning, we aimed at using Robot-based pedagogy. This last falls within the constructivist model of teaching (Di Lieto et al., 2020) and could be applied in many process learning (Augullo et al., 2020). When robots are designed as "objects with which to think", manipulations and experiments based on real situations (Lowrie & Logan, 2006) can take place in a problem-solving environment (Leyzberg D. *et al.*, 2012). In this paradigm, pupils are regarded as "the builders of their knowledge" and of their own intellectual structures (Poletz et al., 2010). They are also epistemologists in the sense that they are led to study their own thoughts in a critical way. Educational robotics may also help to distinguish an error from intellectual punishment. It can give a positive status to the error: programming involves looking for bugs through the analysis, understanding and correcting a program. The error therefore reverts to being a simple step in the process of learning, which could lead to positive behaviors for pupils (Bers, González-González & Armas–Torres, 2019). As the pupils program the robot, they are led to reflection upon their own actions and their own thought processes. In the end, these programmable toys can become true mediation tools wherein pupils identify with the toys through a "mirror effect" (Komis, Misirli, 2011).

However, in the context of French kindergarten, teachers are much more likely to use the body modality in their classrooms. Kindergarten programs require teachers to tackle topics such as the mental construction of numbers, spatial awareness and activities leading to reading. These elements of the program are addressed through the body modality and in the context of recreational learning in particular: during role-playing movement games like "Simon says" or board games like "the Snakes and ladders" or the "Parcheesi game" in which numbers can be used to remember one's position. Some teachers also use games and serious games on tablets (Nacher *et al.*, 2015).

As a way to introduce computer science to five-year-old pupils in their Year-1 (Y 1), the robot seemed to be an in-between modality in which the advantages of the body and of the tablet were combined. To confirm this hypothesis, we thus have to compare the five-year-old pupils' performance scores in decoding, coding and programming design whether they use robot, body or tablet modalities. Indeed, the notion of score is an important aspect of our research work as we are part of the evaluative research approach (Depover, Karsenti & Komis, 2011). Thus, in the context of this paper, our research issue can be summarized as follows: in the framework of an exploring experiment using serious games as mediation with body, robot, and tablet modalities, what types of performance were identified when five-year-old students involved themselves in decoding, coding, and programming design activities in kindergarten classes?

Such a comparative study was made possible by the "*Blue Bot research project*", which first started in 2016; the experiments took place in 2017 with the participation of 230 pupils from 28 classes in the North of France (Bellegarde, Boyaval and Alvarez, 2019). The first results revealed that the robot modality offered better scores than the body or the digital tablet (Alvarez et al., 2019). These first results, however, had been completed on very short deadlines. Therefore, it needed to be confirmed through more advanced statistical processing and analysis and through the exploration of several methodologies. Such work took place from May 2018 to July 2019. These updated results allowed us to shed light on new elements and to provide a more precise analysis of the scores that were linked to the modalities, the latter of which changed according to the types of variables used and on the various cross-analyses. In this paper, we aim **at** present**ing** this statistical endeavor and its ensuing interpretations. To start with, we will introduce the protocol and the nature of the pre-tests and post-tests that were submitted to the 230 pupils in the 28 classes in the North of France. Statistical processing and analysis were done on the data from these tests. We will first focus on the data and then we will interpret and discuss the results. In the context of the discussion, we will see that these analyses seem to confirm the recommendations made by a recent French academic report.

2 Description of the Blue Bot Research Project

This part describes the experimental protocol of the Blue Bot project, then the associated logistics parts.

2.1 The experimental protocol

The Blue Bot project targeted children with the following characteristics:
- Five years of age and enrolled in the Year-1 (Y 1)
- Non-readers
- Laterality under development
- Enumeration under development
- In the pre-reading phase
. In the pre-writing phase: developing handwriting skills.

The *Blue Bot project* is composed of three main dimensions:
- A learning sequence based on play activities
- Educational activities
- Pre-test and post-test evaluations.

Let us now examine these dimensions more closely.

The playful learning sequence

The playful learning sequence aimed to introduce robotics/computer science to five-year-old pupils. It was organized around three time frames: the briefing, the event, and the debriefing. Such sequencing was based on the "three educational time periods" model defined by Nicole Tremblay (Tremblay, 2007). During the briefing phase, the teacher tells a short story in which the robot is the main character and must go on a journey. Pupils are asked to transcribe the trajectory: it is composed of a starting point, a final point, and it follows a predetermined progression. Concerning cognitive mediation, we used three variations of the same game principle involving a robot. This was done in an attempt to introduce robotics/computer science to the pupils, and to coding in particular. These three variations differed due to the nature of their modalities:

- **Use of the body**: a pupil embodies the robot and must move on a checkerboard marked on the floor. Other pupils give him/her instructions (Figure 1 – left image).
- **Use of the robot:** the pupils program the Blue Bot toy robot as it travels on a checkerboard that is printed on a plastic rug and placed on a table (Figure 1 – center image)..
- **Use of the digital tablet:** the game is identically replicated in a completely virtual environment. It is played on a tablet (Figure 1 – right image).

Figure 1. *From left to right, the three modalities used for the serious game: body, robot and tablet*

Regardless of the modality (body, robot, tablet), the programming activities always took place on a 24 square checkerboard (4 by 6 grid); the aim was always to move the robot from an initial starting point to an end point. And then a progression occurred from level to level. As illustrated in Figure 2, we included obstacles to avoid, such as a growling dog, before the robot could reach the end point (the friendly robot).

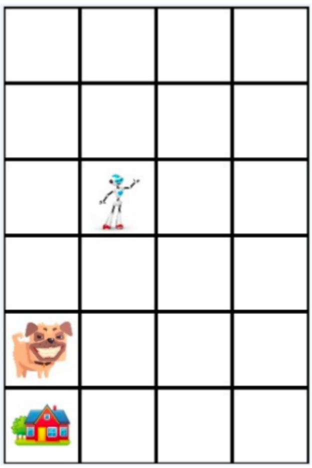


Figure 2. *Example of a level proposed for a learning activity within the Blue Bot project experiment*

The educational activities

In order to prepare the pupils to play the various variations of the game, we designed a variety of educational activities, detailed below.

- ***Reading an excerpt from the* Vibot story 5** (Romero & Loufane, 2016) : the aim of this discovery activity was, on the one hand, to raise the pupil's interest for the robot object, and on the other hand, to make sure that he/she would get involved in the subsequent programming activities. The robot was the central character in all of these experiments, regardless of the type of cognitive mediation.

- **The four time periods focusing on teaching the algorithmic element.**

The teaching components dedicated to algorithmics were offered to the pupils once the preliminary activities had been completed. These teaching units were structured in four stages:

- ***1st stage: being introduced to algorithmic and programming instructions*** – how to handle the robot, the tablet, the checkerboard and understand the properties of the numerous controls (on the robot, on the tablet, from the body and through the directional pictograms)
- ***2nd stage: being progressively introduced to the Blue Bot robot's controls*** (whatever the modality – body, robot or tablet)
- ***3rd stage: creating a coding sequence*** by using a programming bar system for each type of teaching/learning situation- A problem situation suggested by the teacher: driving the robot to a precise place.
- ***4th stage: creating a coding sequence*** using the programming bar system with the addition of more constraints (fixed obstacle, path to follow, etc.) and problem-based situations suggested by the teacher based on these constraints.

The pre-test and post-test evaluations

The pre-test and post-test activities were meant to assess the pupils' performances in the field of programming. These activities were structured into 3 sections:

Activity #1 - decoding: decoding a set of instructions that needed to be depicted as a traced path on a grid. The pupil was asked to read the various instructions represented by arrows and to trace the Blue Bot robot's trajectory on the grid (Figure 3).

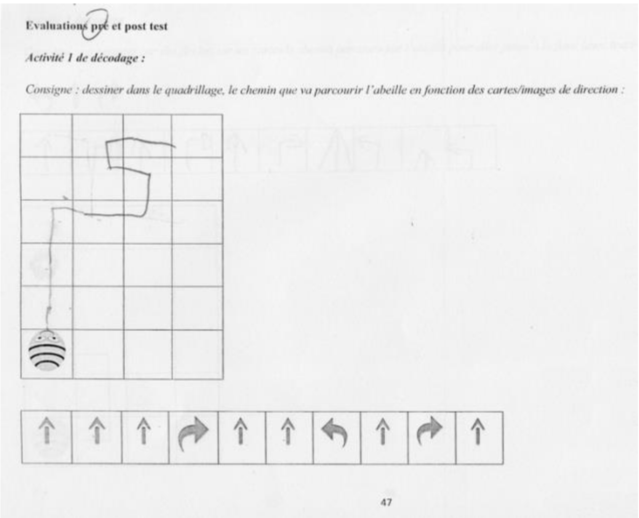


Figure 3. *Decoding activity 1*

Activity #2 - coding: coding a trajectory based on a given layout on the grid. The pupil was asked to deduce and transcribe the list of instructions thanks to arrows, based on the suggested itinerary drawn on the grid (Figure 4).

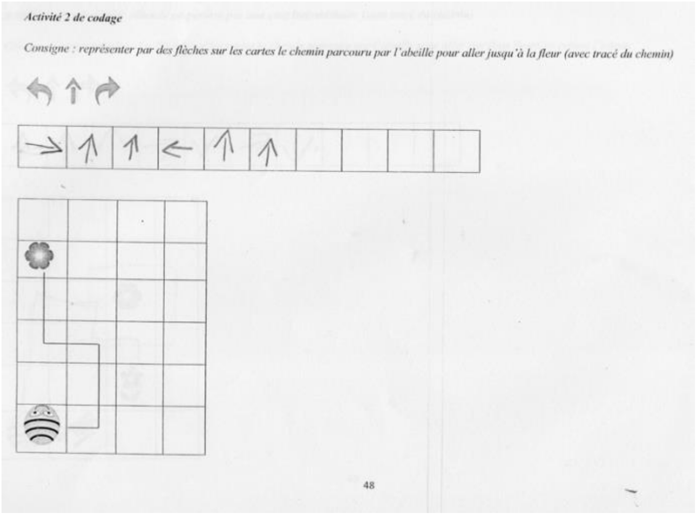


Figure 4. *Coding activity 2*

Activity #3 – programming design: presenting a path which respected the associated constraints (elements that the robot had to reach and others that had to be avoided) with an associated coding proposition. The pupil was asked to determine a path for the robot by tracing it on the grid. Using arrows, the pupil then had to submit the instructions associated with his/her path. When the pupil was tracing the path, he/she had to take into account the robot's wish to pass by two flowers and to avoid one bird. Two paths were possible: the most optimized one was coded with seven instructions whereas the less optimized one was coded with nine instructions (see Figure 5).

To evaluate the different activities, a scoring system has been set up:

- **Activity #1 - decoding evaluation**
  - **Number of correct instructions** (Score out of 10 points): How many instructions out of the 10 proposed are well transcribed on the grid?
  - **Understanding instructions** (Score out of 1 point): Does the pupil reproduce the movement of the Blue Bot robot on the grid?
  - **Understanding principle of robot rotation** (Score out of 3 points): Out of the 3 expected rotations, how many times does the pupil spin the robot on the spot without moving it?
  - **Reaching the final square** (Score out of 1 point): Does the path proposed by a pupil actually end on the grid's final square, as expected?
  - **Good counting** (Score out of 7 points): Does a pupil copy the good number of steps corresponding to the arrows facing forward on the grid? (by group of arrows)

- **Good L/R laterality** (Score out of 3 points): For each rotation, does a pupil put the robot in the right direction? (3 rotations in total)

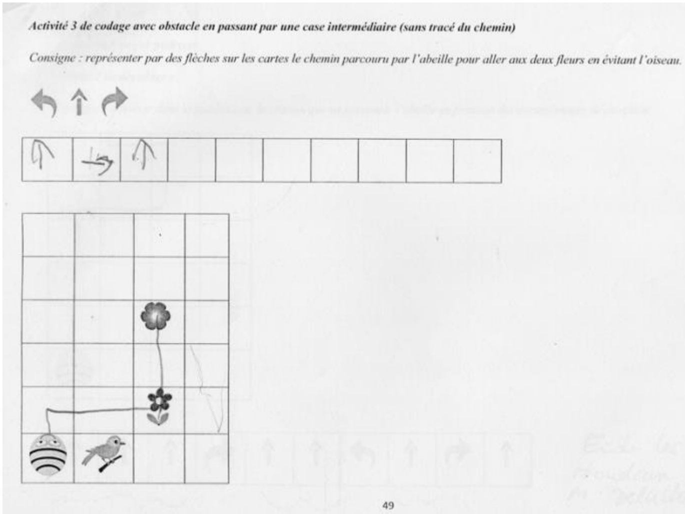


Figure 5 *Programming design activity 3*

- **Activity #2 - coding evaluation**
  - **Number of correct instructions** (Score out of 10 points): How many instructions out of the 10 proposed are well transcribed on the grid?
  - **Number of instructions used** (Score out of 10 points): How many of the 10 instructions are used on the grid?
  - **Understanding principle of robot rotation** (Score out of 4 points): For the 4 expected rotations, is a pupil capable of rotating the robot at the right time?
  - **Reaching the final square** (Score out of 1 point): Does the path proposed by a pupil actually end on the grid's final square, as expected?

- **Activity #3 - programming design evaluation**
  - **Number of instructions used** (Score out of 10 points): How many instructions are proposed by the pupil?
  - **Number of correct instructions** (Score out of 10 points): How many correct instructions are proposed (7 or 9)?
  - **Optimized way** (Score out of 1 point): Does the pupil propose the optimum way with 7 instructions?
  - **Non optimized way** (Score out of 1 point): Does the pupil propose the non optimum way with 9 instructions?
  - **Understanding bird instruction** (Score out of 1 point) : Does the pupil avoid the bird?
  - **Understanding flowers instruction** (Score out of 2 points) : How many flowers does the pupil propose to reach (0, 1 or 2)?

2.2 Logistical aspects

Since the number of available robots and tablets was limited, we planned a four-week rotation to allow all the classes to participate in the experiment. These rotations enabled us to expose all of the classes to the different modalities (robot, tablet, body) at different stages and in different sequences, all of which served the experimental protocol. We used this rotation to our advantage to elaborate a variety of combinations of modalities (body, robot, tablet) as shown in Table 1. The post-test phases were placed at different stages so that seven different combinations of modalities could take place:

- #1 Body alone (B)
- #2 Robot alone (R)
- #3 Tablet alone (T)
- #4 Body + Robot (BR)
- #5 Body + Tablet (BT)
. #6 Robot + Tablet (RT)
- #7 Body + Robot + Tablet (BRT)

An eighth "Placebo" group (P) was created to conduct pre-test and post-test evaluations without any educational activities in between. This was done to check if the teaching areas in the seven other groups (B, R, T, BR, BT, RT and BRT) were truly connected to several educational activities and serious games on offer.

B (4 classrooms)	**R** (4 classrooms)	**T** (4 classrooms)	**BR** (4 classrooms)	**BT** (4 classrooms)	**RT** (4 classrooms)	**BRT** (4 classrooms)	**P** (5 classrooms)
pre-test	**pre-test**	**pre-test**	**pre-test**	**pre-test**	**pre-test**	**pre-test**	**pre-test**
body	**robot**	**tablet**	**Body**	**body**	**robot**	**body**	**-**
post-test	**post-test**	**post-test**	**Robot**	**tablet**	**tablet**	**robot**	**-**
robot	**tablet**	**body**	**post-test**	**post-test**	**post-test**	**tablet**	**-**
tablet	**body**	**robot**	**Robot**	**tablet**	**body**	**post-test**	**post-test**

Table 1. *Blue Bot research groups and calendar*

The various pre-tests and post-tests were organized by the teachers on scheduled dates. Once the pupils had completed the documents (see Figures 3, 4 & 5), the teachers provided the researchers with material, by post or by hand. The researchers assessed the pre- and post-tests themselves to make sure that the tests were processed homogeneously. In addition, the researchers in charge of assessing the tests did not conduct experiments in the different schools. Therefore, they could not establish a link between the pupils and the tests; this ensured a certain level of neutrality. Researchers discussed the evaluations together, in case the tests would show certain anomalies or if some answers would have been veered away from the usual standards. Once these evaluations were completed, all of the data was statistically processed.

3 Methodology

3.1 Processing and sorting the data

Finding usable data

Out of the 28 classes taking part into the experiment, there were 255 pre-tests in total. This amounted to 255 observations, since one observation corresponded to one pupil in our data set. However, some pupils had been absent during the pre-test or the post-test phase. We didn't take into consideration the data associated with these pupils since the progress of the performances between both stages could not be assessed. After we had removed these missing pupils, there were 243 pupils left.

We should specify that each of these 243 pupils was characterized by the questionnaire answers (pre-tests and post-tests), by the modality group to which they belonged (B, R, T, BT, BR, BT, BRT and P) and finally, by their class, the latest being represented by their teacher's name. All pupils and teachers were anonymous. We modified all names appearing in the results to guarantee the anonymity of the participants.

Not only did some teachers from various classes fail to send the pre-tests and post-tests, but some documents also got lost. Due to unavailable data, the BT (Body and Tablet) group could not be studied. This reduced the number of groups to seven: B, R, T, BT, BR, BRT and P.

List of variables associated with the data

The list of variables was based on the various pre-test and post-test assessment criteria. The number of variables for the pre-test and post-test phases was the same since the activities and assessment criteria were identical. The decoding activity was therefore linked to 14 variables: 7 in the pre-test phase and 7 in the post-test phase. The coding activity was composed of 8 variables: 4 in the pre-test phase and 4 in the post-test phase. Finally, there were 12 variables for the design activity: 6 in the pre-test phase and 6 in the post-test phase. The total number of identified variables was 34: some were numeric and only two of them were nominal. Numeric variables referred to formal references that could be counted objectively: number of correct instructions, accurate count of rotations that the robot had to perform during the trajectory.

In statistics science, a variable is called numeric when its values are numbers with or without a unit thus arithmetic operations can be conducted (+ - x ÷). Note that Boolean variables (Yes/No) have been coded as numeric variables.

The nominal variables were related to non formal criteria which could give way to subjective interpretations, such as the pupil's overall understanding of the instructions or a student's ability to handle laterality.

Table 2 presents the list of all the variables used for the statistical processing of the data relating to the pupils' performances when learning computer science in the context of the Blue Bot project. The type of activity relating to each variable (decoding, coding, programming design) is specified in the table. Also specified is the name of the variable and whether it corresponds to the pre-test or the post-test phase. In the following, only numeric variables are taken into account to develop the tests.

Type of activity	**Name of the pre-test variable**	**Name of the post-test variable**	**The variable's possible values**	**Type of variabl e**	**Explanatory description**
Decoding	Act1aV1	Act1bV1	0 to 10	Numeric	Number of correct instructions: Out of the 10 instructions, how many have been correctly transcribed onto the grid?
Decoding	Act1aV2	Act1bV2	0 or 1	Numeric	Understanding the instruction: did the child reproduce the Blue Bot's itinerary onto the grid?
Decoding	Act1aV3	Act1bV3	0 to 3	Nominal	Understanding the static rotation: Out of the 3 rotations presented, how many times does the child rotate the robot statically without making it move forward?
Decoding	Act1aV4	Act1bV4	0 or 1	Numeric	Reaching the final square: does the child suggest an itinerary that ends on the expected final square on the grid?
Decoding	Act1aV5	Act1bV5	0 to 7	Numeric	Accurate count: Does the child accurately transcribe to the grid the number of steps that go with the arrows pointing forward? (by group of arrows)
Decoding	Act1aV6	Act1bV6	0 to 3	Numeric	Good L/R laterality: For each rotation, does the child position the robot in the right direction?
Decoding	Act1aV7	Act1bV7	0, 1 or 2	Nominal	Type of representation (Instructions: 1 / Tracing: 2 / Other: 0): Does the child represent the trajectory through a series of arrows (Instructions) or through a continuous line (Tracing) or in another way?
Coding	Act2aV8	Act2bV8	0 to 10	Numeric	Number of correct instructions (/10): Out of the ten instructions, how many were accurately transcribed onto the grid?
Coding	Act2aV9	Act2bV9	0 to 10	Numeric	Number of instructions used: Out of the 10 instructions, how many were used on the grid?
Coding	Act2aV10	Act2bV10	0 to 4	Numeric	Understanding the static rotation: Out of the 4 expected rotations, does the child suggest turning the robot at the right time?
Coding	Act2aV11	Act2bV11	0 or 1	Numeric	Reaching the final square: does the child suggest an itinerary that ends on the expected final square on the grid?

Design	Act3aV12	Act3bV12	0 to 9	Numeric	Number of correct suggested instructions: How many instructions does the child suggest, either for the optimized trajectory (7 instructions) or for the longer path (9 instructions)?
Design	Act3aV13	Act3bV13	0 to 10	Numeric	Number of instructions used: how many instructions did the child use?
Design	Act3aV14	Act3bV14	0 or 1	Numeric	Optimized trajectory: Did the child suggest using the optimized trajectory with 7 instructions?
Design	Act3aV15	Act3bV15	0 or 1	Nominal	Long trajectory: Did the child suggest using the longer trajectory with 9 instructions?
Design	Act3aV16	Act3bV16	0 or 1	Numeric	Understanding the bird rule: Did the child avoid the bird?
Design	Act3aV17	Act3bV17	0, 1 or 2	Numeric	Understanding the flower rule: how many flowers does the child suggest reaching? (0, 1 or 2)

Table 2. *Presentation of all variables used for the statistical processing of the Blue Bot project's performance study*

The nomenclature associated with the variables

The names of the variables recorded in Table 3 correspond to a nomenclature defined as follows: "Act" for "activity", followed by a number (1=decoding, 2=coding, 3= design), followed by a letter (a= pre-test phase, b=post-test phase), the letter "V" for "variable" followed by a number ranging from 1 to 17, standing for the variable's number. Therefore, "Act1bV7" refers to a variable for a decoding activity at the post-test stage and is the seventh variable on the list.

Deleting an anomaly

All of the variables and ensuing values were entered into a spreadsheet. During this operation, an anomaly was detected for one of the school classes in the post-test phase. All of the 13 pupils in this particular class had obtained the maximum score in some coding, decoding and programming design activities. Such data seemed highly implausible when compared with the heterogeneous figures collected in all of the other classes. Therefore, we chose not to include the pre-tests and post-tests for this class in our data processing. This lowered the number of pupils for whom we could use the data from 243 to 230.

Deleting three variables

Following discussions between researchers and statisticians, three variables were disregarded out of the 17 variables listed in Table 3: V3 (Understanding the static rotation), V7 (Type of representation) and V15 (Long trajectory) due to the fact that similar variables did not exist in the design activity. This brought the total number of variables analyzed in this experiment down to 15.

3.2 Evaluating the performances of the P Group (Placebo)

The P group (Placebo) allows us to appraise the educational efficiency of the project. Indeed, this group received neither training in computer programming nor in robotics at school between the pre-test and post-test phases. Did this group show significantly lower performance rates than the other groups? The results needed to be conclusive for us to deduce whether the experiment was reliable or not. To answer this question, it is necessary to conduct a statistical test. Beforehand, we have built a variable named "S" that assesses the difference between the P group and the other ones.

Construction of "S"

The variable named "S" quantifies the number of points earned between the pre-test /post-test phases. To evaluate S, we only chose the 6 numeric variables from Table 3.

S = (Act1bV1 - Act1aV1) + (Act1bV5 - Act1aV5) + (Act1bV8 - Act1aV8) + (Act1bV9 - Act1aV9) + (Act1bV12 - Act1aV12) + (Act1bV13 - Act1aV13)

S was then rendered into a grade out of 20. Negative grades were possible. Grades fit in the [-7.2 ... 16.4] range. We identified 4 classes:

- Negative grade : Reg
- Grade [0.. 5[: Low
- Grade [6..10[: Average
- Grade >10. : Honorable

At this stage, each pupil was characterized by 2 variables: :

- by his/her coded grade following one of the 4 classes (Reg / Low / Average / Honorable)
- and by his/her group, either known as control (written as P for Placebo) or as other (written as ACTIVE).

In Table 3, we crossed these two variables for the total population of 230 pupils.

Group/Grade	Honorable	Average	Low	Reg	Total
ACTIVE	7	43	104	29	183
P	0	3	31	13	47
Total	7	46	135	42	230

Table 3. *Crossing grades and affiliation with the control group*

Khi2 test

Table 3 indicates that P group have a score of 31 Low and 13 Reg, i.e. 97% of the pupils of this group. To confirm this dependence, we carry out a statistical test of the crosstab.

	Khi2	%khi2	DIFFERENCE	%cumulated Khi2
P/Honorable-Average-Grade	5.7	53.9	NEGATIVE	53.9
P/Reg-Grade	2.3	21.6	POSITIVE	75.5
ACTIVE/Honorable-Average-Grade	1.5	13.8	POSITIVE	89.4
ACTIVE/Reg-Grade	0.6	5 .6	NEGATIVE	94.9
P/Low-Grade	0. 4	4.0	POSITIVE	99.0
ACTIVE/Low-Grade	0.1	1.0	NEGATIVE	100
Total	10.5	100%		

Table 4. *Calculation of Khi2 from the data of Table 3, sorted in descending order according to the % of KHI2*

We carry out a step-by-step analysis of khi2 (Cibois, 1987). Due to the low number of honorable grades (7), honorable and average grades were consolidated into one single class known as "Honorable-Average-Grade". We then proceeded with the calculations recorded in Table 4.

Three observations emerge when reading Table 4: :

- 1) The calculated Khi2 is 10.5. It is therefore significant (theoretical Khi2 is 5.99 p=5% ddl=2)
- 2) 94.9 % of the Khi2 is due to the absence of acceptable grades in the P group and of an overrepresentation of Reg-Grade in the P group
- 3) The opposite effect can be observed for the active group: an absence of Reg-Grade in the active group and a strong presence of Honorable-Average grades in the active group.

We can conclude that the P control group presents grades that are significantly (at a risk of 5%) lower than those of the other groups.

3.3 Constructing a set of variables to measuring performances according to the modalities

The Table 3 refers to the 230 pupils who were there at the beginning of the experiment: the P group represents 27 pupils, the Active group gathers 183 pupils (disaggregated in 6 modalities B, BR, BRT, R, RT, T). We must also notice that the group combining the three modalities – Body, Robot and Tablet (BRT) – was only composed of 6 pupils. Due to the small number of participants related to the BRT group, we therefore chose to disregard this one, for the remainder of the statistical processing.

The following analyses feature a headcount of 177 pupils divided into 5 groups (See Table 5). These 177 pupils can be considered as a sample: the theory of probabilities makes it possible to indicate whether two percentages are significantly different (See paragraph 3).

Groups	**Headcounts**	**Percentage in viewof the overall headcount (177 pupils)**
B	26	14.7%
BR	45	25.4%
R	28	15.8%
RT	60	33.9%
T	18	10.2%
Total	177	100%

Table 5. *Headcounts for each modality group out of the 177 remaining pupils in the active group*

Now, the objective is to analyze the performance of these 177 students divided into five groups (B, R, T, BT and BR) for the three main activities (decoding, coding and design). To measure the performances, we created 3 numeric variables: Decoding-Score, Coding-Score and Design-Score:

Decoding-Score = (Act1bV1 + Act1bV2 + Act1bV4 + Act1bV5 + Act1bV6) - (Act1aV1 + Act1aV2 + Act1aV4 + Act1aV5 + Act1aV6)

Coding-Score = (Act2bV8 + Act2bV9 + Act2bV10 + Act2bV11) - (Act2aV8 + Act2aV9 + Act2aV10 + Act2aV11)

Design-Score= (Act3bV12 + Act3bV13 + Act3bV14 + Act3bV16 + Act3bV17) - (Act3aV12 + Act3aV13 + Act3aV14 + Act3aV16 + Act3aV17)

These three scores were standardized as a grade ranging between - 20 and + 20. This grade was then re-coded with the honorable, average, low and reg classes using the same table as done previously (see 2.2.). We obtained a new "grade" variable composed of four items: honorable, average, low, reg. Each pupil was therefore characterized by his/her group (see Table 5) and by a "grade" for each for the 3 main activities.

We crossed the "grade" variable with the "group belonging" variable to see whether there was a dependency link between the groups and the grade classes. We did this for the three activities.

To analyze this statistical dependence, we created six histograms in the following way:

- Analyzing the line percentages which, in our case, reflected a grade modality distribution in the five groups (see Figures 6, 8 and 10).
- Analyzing the column percentages which, in our case, reflected the distribution of the different groups for one grade modality (see Figures 7, 9 and 11).

To summarize the most significant results, it was necessary to recode the grade variable according to two grades:

- Honorable - Average
- Low - Reg

This recording makes it possible to significantly differentiate the 3 main activities (see Figure 12). It also allows a ranking of the modalities by assigning them scores (see Figure 13). Finally, these results lead us to divide the modalities (see Figure 14) into two classes:

- Mono-modalities (B, R, T)
- Bi-modalities (BR, RT).

4 Results

The results are presented in two steps. A micro step where each activity is viewed through the different modalities of the groups. A second macro step which aims to synthesize the most significant results.

In the macro step we use probabilities to calculate the error margin. We call on the following statistical method: given an infinite population where a proportion p of individuals have a certain characteristic, the method consists in finding a confidence interval of p from a proportion f observed in a sample of size N. In statistics (Saporta, 2006, p.211), we show that:

$$f - t_\alpha \cdot \sqrt{\frac{f(1-f)}{N}} < p < f + t_\alpha \cdot \sqrt{\frac{f(1-f)}{N}}$$

For *Activity performance* (see 3.2.1), we have N=177. We have fixed the risk at 5%, i.e.t_α= 1.96 so we have an approximate error margin of 7%.

For *Modalities assessment* (see 3.2.2), the number of cumulated modalities for Honorable-Average grade, whatever the activity, is 266. The number of cumulated modalities for Low-Reg grade, whatever the activity, is 265. So we can take 265 for N and with the same risk we have a error margin of about 6%.

For the micro-step, the number of individuals in each modality is too small to use this formula.

4.1 Analysis of each group's performance for the three activities

We examined performances according to the activities. Thanks to the histograms, we could then analyze the performances of the B, R, T, BT and BR groups in keeping with the three decoding, coding and programming design activities.

Examining the decoding activity

The Tablet group (T) is isolated from the others. At least two elements confirm this position: the highest number of negative grades (25% - see Figure 6) are found in this group and 50% (see Figure 7) of the group displayed negative or low grades.

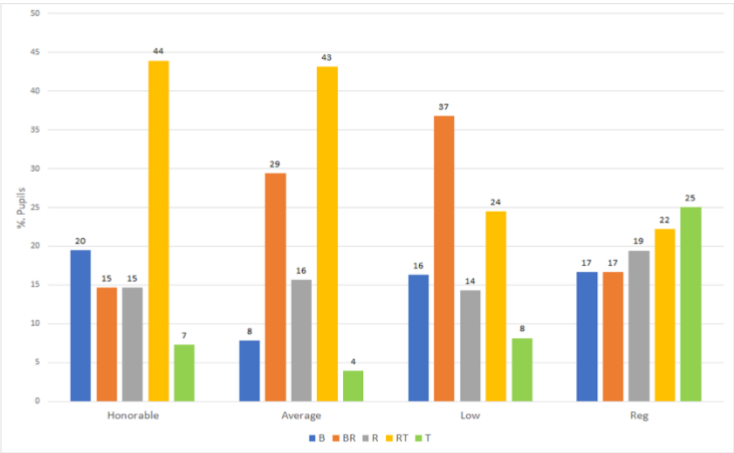


Figure 6. *Decoding activity: distribution of the inter-group grades*

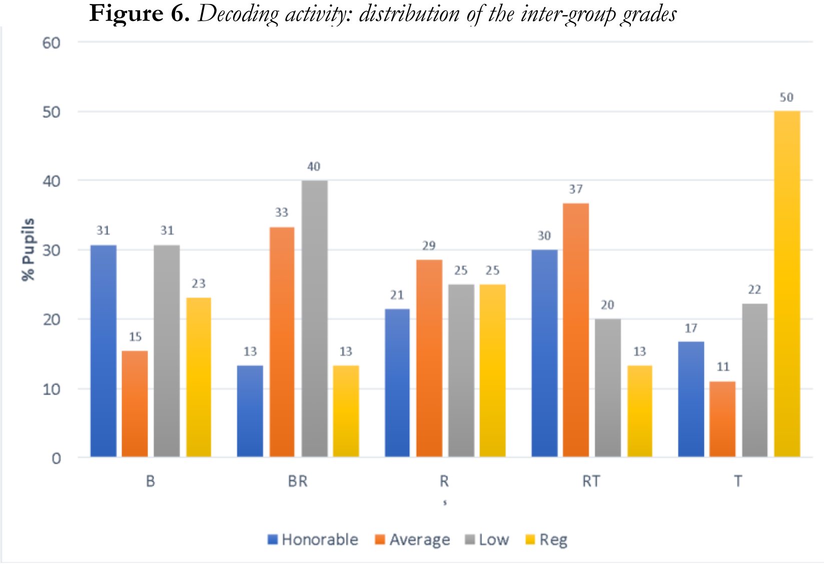


Figure 7. *Decoding activity: distribution of the intra-group grades*

The Figure 7 also shows that the closest groups are Robot (R) and Body (B). It should also be noted that a flat histogram can be observed for the R modality (21%-29%), in contrast with the T modality in which there are larger differences (11%-50%) between scores.

Examining the coding activity

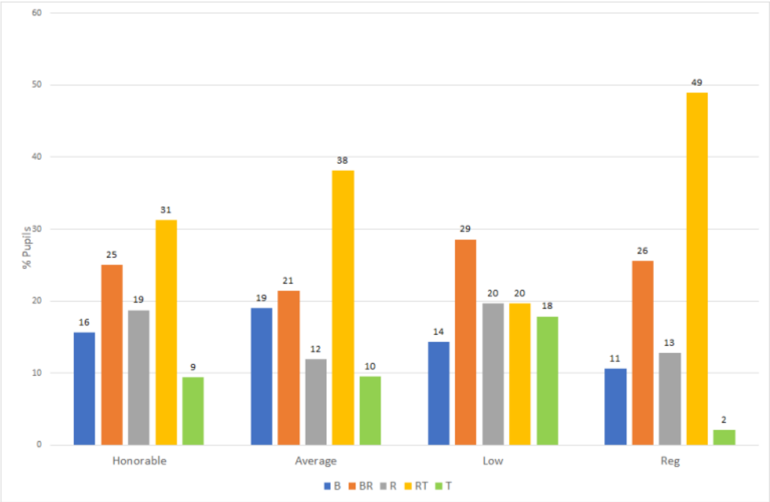


Figure 8. *Coding activity: distribution of the inter-group grades*

The Robot-Tablet group (RT) is isolated. At least two elements confirm this position: first, the highest number of negative grades (49% - see Figure 8) are found in this group; second, 38% of the group (see Figure 9) ended with a negative grade.

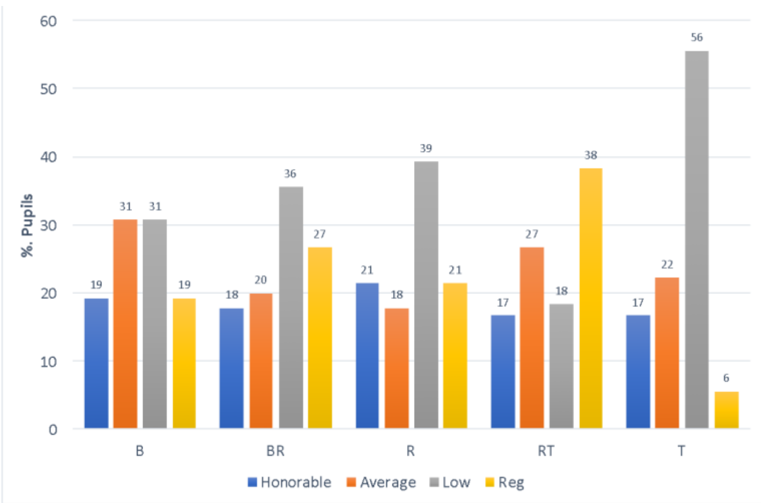


Figure 9. *Coding activity: distribution of the intra-group grades*

It also asserts that the closest groups are the Robot (R) and the Body-Robot (BR) (see Figure 9).

Examining the programming design activity

Figure 10 shows that the intensity of the Honorable grade (54%) of the RT modality is much higher than the other modalities.

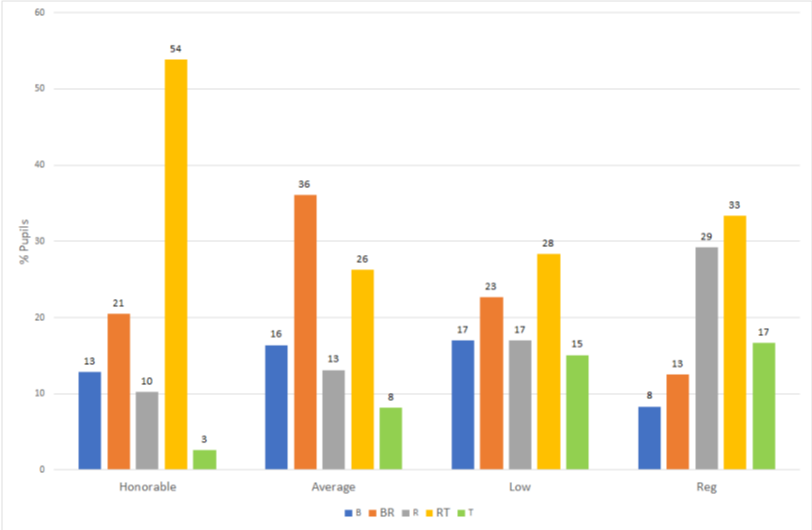


Figure 10. *Programming design activity: distribution of the inter-group grades*

We can see that a vicinity exists between the Body group (B) and the Body Robot group (BR) (see Figure 11). A second vicinity of the same intensity between Robot (R) and Tablet (T) can also be seen in Figure 11.

To sum up, we can note that, for the three activities, the Robot Tablet (RT) group presented the highest amount of honorable grades: 44% for decoding (See Figure 6), 31% for coding (See Figure 8), 54% for programming design (See Figure 10). The highest number of pupils who performed well could be found in this group.

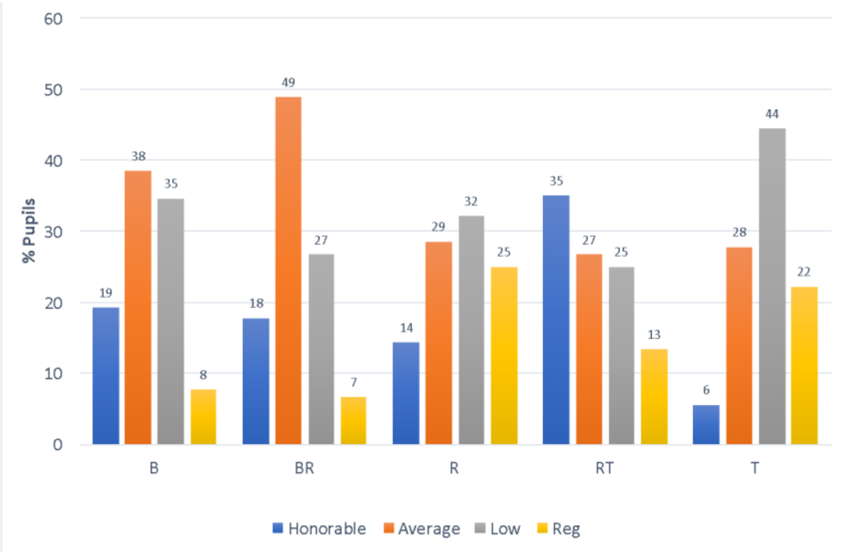


Figure 11. *Programming design activity: distribution of the intra-group grades*

4.2 Summary of results

From the previous findings, we will summarize the main lessons of this experimentation concerning the 3 types of main activities and the different modalities.

Activity performance

The grades of all the 177 pupils were put together. We obtained these grades by calculating the difference between the post-test scores and the pre-test scores, all modalities combined (B, R, BR and RT). The results appear in percentages. This approach allowed us to identify how the performances – represented by grades – had evolved for each activity. Figure 12 shows a representation of the headcount distribution according to the various developments of the grades crossed with the different activities.

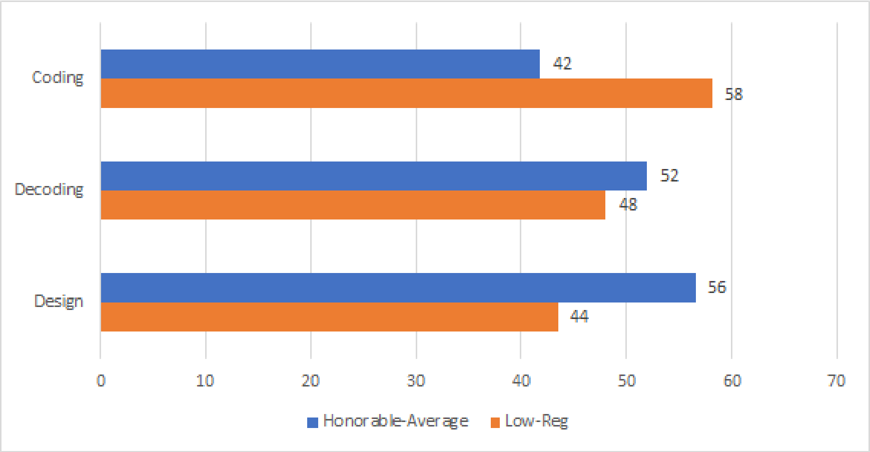


Figure 12. *Cumulated grades for all modalities according each activity*

Figure 12 shows that Y 1 produced the most Honorable and Average grades (56 % with a margin of error of 7 %) when they engaged in the programming design activity. The decoding activity shares approximately the same rate of honorable grades and low Reg.. On the contrary, the number of negative ratings for the coding activity is significantly much higher (58% with the margin of error of 7 %) than the one of the program design activity (44%).

Modalities assessment

The grades of the modalities are gathered whatever the activities. The results are visible in percentages.

Figure 13 shows a strong dominance of the RT and BR modalities. They reach 65% (with a margin of error of 6 %) of the Honorable-Average grade whereas the three modalities B, R and T accumulate 47% (with a margin of error of 6 %) of the Low-Reg grades. Finally, we notice the very low score of the tablet modality.

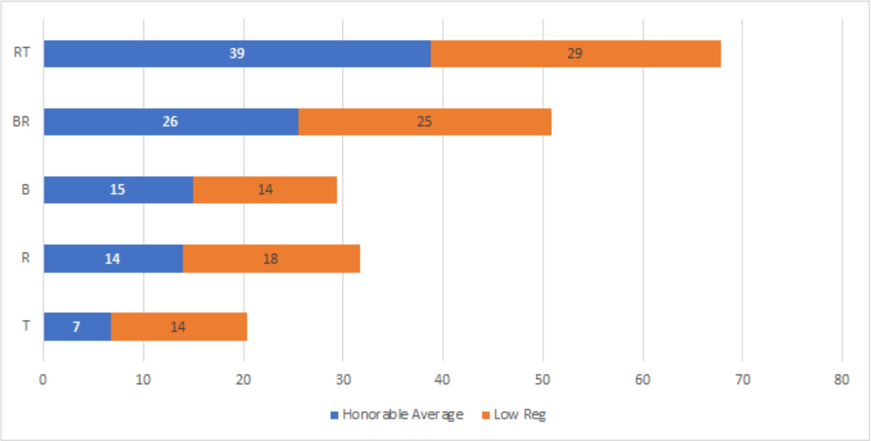


Figure 13. *Accumulated modalities for each grade whatever the activity*

Mono-Modality versus Bi-Modality assessment

The previous results lead us to merge the Mono-Modality groups (B, R and T) and those related to the Bi-Modalities groups (BR and RT). The grades of these two main groups are gathered for the 3 main activities. We then obtain Figure 14.

This shows a preeminence of the group of Bi-Modalities which accumulate 64% of the Honorable-Average grades as opposed to the group of Mono-Modalities which accumulate 45% of the Low-Reg grades.

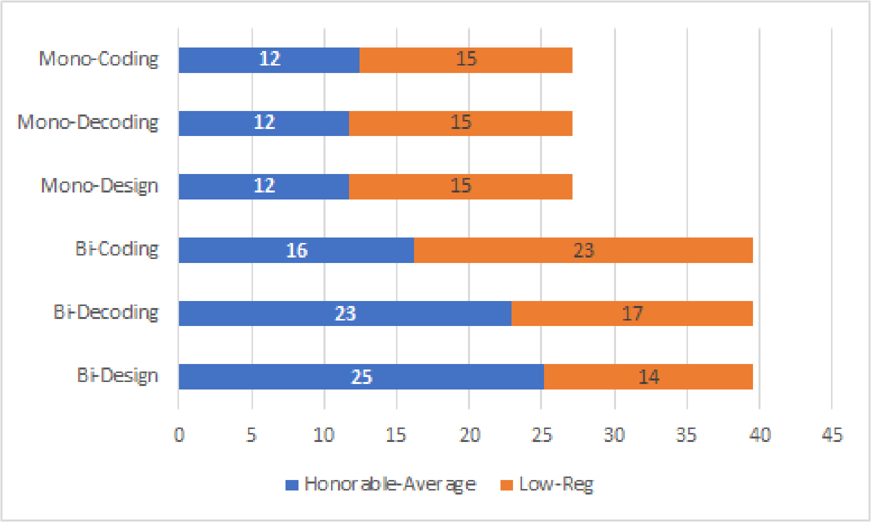


Figure 14. *The performances of mono-modalities versus bi-modalities*

Now that all these aspects of the experiment have been examined, we can interpret and discuss the results.

5 Discussion

The headcount for each modality group is too small and the micro analysis does not allow definitive conclusions to be drawn from the statistical result. We must remain cautious. Nevertheless, the macro study allows us to identify trends.

5.1 Comparative analysis of the modalities according to the different activities

The teachers requested that the body modality be compared with the robot and the tablet. As discussed above, teachers often use the body modality in their teaching practices (see 1.3). We therefore established the Body (B) modality as a reference base. To satisfy the teachers' request, we needed to focus on the modalities that offered higher performance rates than B.

A Review of the modalities with higher performance rates than Body (B)

Figure 13 shows us:

- RT and BR modalities have a higher rate of Honorable-Average grades than B
- RT gets the higher rate of Honorable-Average grades
- T presents the lowest rate of Honorable-Average grades

In summary when we focus on the Honorable-Average grades, we can observe that all modalities show higher scores than B apart from T.

For the coding and programming design activities, Figures 8 and 10 show us that the scores of the BR and RT modalities are higher than B.

At this stage, some modalities seem to yield higher performances than B. More precisely, the RT modality is more effective than the B one in all activities. The BR modality effectiveness was also higher than the B one, in all cases, except for the honorable grades in the decoding activity (see Figure 6).

Accounting for the regression grades

While our focus is on the positive performances, examining the regression grades is nevertheless useful to complete this review. A regression means that the Y 1 performed better during the pre-tests than during the post-tests.

Figure 6 shows regression rates of B, BR and R for the decoding activity. These ones are almost identical. Only T has a significantly higher regression rate than B.

For the coding activity, T is the only modality that is lower than B in terms of regression (see Figure 8). For the Design activity, we can notice that the B modality shows the lowest regression score (see Figure 10).

The modalities B and R are clearly comparable. But we can state that for all the activities combined, the modalities RT and BR are more prevailing than modality B (figure 13).

Therefore, we can conclude that the B modality presents the lowest regression rates for all combined activities.

The paradox of the Robot-Tablet (RT) modality

For both the Decoding process/activity, and the Design one, the RT modality seems to exceed the others. However, it doesn't apply to the coding exercise. This should be analysed with nuances since coding is the only activity from all modalities which includes a Low-Reg rate quite superior than its Honorable-Average grades. Last but no least, if we add up activities, Figure 13 states that the RT modality holds a much higher H-M grade than all the others.

However, when we focus on the R and T modalities separately (see Figure 13), we observe that their performances are not higher than the ones of the B modality. It seems surprising to see that the association of two modalities would produce a higher-performing pair when these modalities are not necessarily more efficient when taken separately. How could this seemingly paradoxical phenomenon be explained? Another question emerges as well: why is the association of T and R more effective than the association of B and R? From the symposium discussions between researchers and teachers, one hypothesis emerged: the "double-modality" hypothesis. Let us now describe it in detail.

5.2 The double-modality hypothesis

The double-modality hypothesis is based on the notion that it is more effective to experiment with two different modalities rather than only one.

Confirming the hypothesis by analyzing the BR bi-modality

We used the RT bi-modality to substantiate the double-modality hypothesis since it showed the best performances among the pupils tested in the decoding, coding and programming design activities. When we look at Figure 13, we also see that BR also generated good performances. While the grades associated with BR are lower than RT, the majority of them are greater than B, R and T as individual modalities. If we were to rank the modalities according to the accumulated honorable and average grades, RT would be the first, followed by BR, and then followed by B, R, to finish by T. At this stage, the double-modality hypothesis therefore seems consolidated.

However, we must note that the RT bi-modality is also very divisive since it generated a high number of regression grades (see Figure 13). In this respect, the BR bi-modality seems to present a high regression rate for the coding activity. It is in second place after RT (see Figure 8). However, for the design programming activity, its regression rates are barely higher than **for** the B modality, the latter of which is the least regressive (see Figure 10). As for the decoding activity, the BR modality is equal to B in terms of regression scores (see Figure 6). Therefore, we can deduce that the high regression rate is mainly due to the RT bi-modality and that it cannot necessarily be generalized to all other bi-modalities. Let us try to verify this.

Comparing the performances of the mono-modalities and of the bi-modalities

To check the double-modality hypothesis, we must make sure that, in light of the mono-modalities (B, R and T), the Y 1 increase their performances more than they regress when using the bi-modalities (BR and RT). With this in mind, we can refer to Figure 14. This presents the performances of mono-modalities versus bi-modalities and clearly shows that when the honorable and average grades are accumulated, the bi-modality outpaces the mono-modality.

Let us now analyze the regressions. In the case of the coding activity, Figure 14 shows that the bi-modalities present very high regression rates (REG) in comparison with the mono-modalities except for the programming design activities. But how can we explain this phenomenon?

The bi-modalities, a question of time

The time factor may play a preponderant role when attempting to explain the differences between the performances of the bi-modalities and of the mono-modalities. The pupils who experimented with two modalities spent twice as much time on serious gaming as the pupils who only explored one modality. Therefore, the additional time may have served to promote learning in the case of the bi-modalities. However, the notion of time does not explain why the bi-modalities presented such a high regression in the case of the coding activity. This statement solely uses the time factor to explain the impact of the bi-modalities. Other factors, which still need to be identified, are necessarily involved and have a positive or negative influence.

5.3 Understanding the results of the coding activity compared with the programming design Activity

As seen above, the double modality gives rise to many questions. For instance, how can we explain the fact that the statistics indicate better performance rates for the programming design activity than for the coding activity (see Figures 12 & 14)? Thus, other hypotheses need to be found.

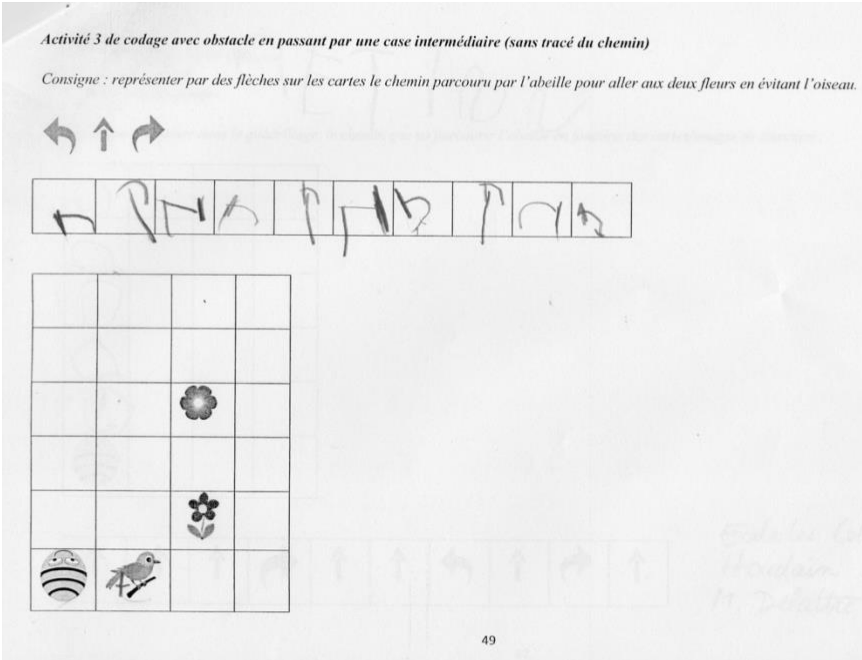
Two possibilities were discussed during the symposium organized by ourselves, gathering the researchers, teachers and statisticians.

The first one questions the nature of the pre-tests and of the post-tests. Were they suited to Y 1 ? Could it be that some of the pupils may have been confused by the modality and the instructions of the pre- and post-tests? For example, the five-year-old pupils were asked to draw arrows to represent commands during the pre-tests and the post-tests for the coding activity (see Figure 4). But this may require teaching them specific skills to represent such symbols when indicating orientations: left, right, or above. Since the pupils had not been taught those skills, they may not have been able to produce the appropriate arrows to complete the coding activity. This hypothesis is confirmed by some documents produced by different pupils during the tests as shown in Figure 15. We can see that it is more difficult for some pupils (see Figure 15 up) to draw arrows than for others (see Figure 15 down).

Therefore, this first idea seems, at first glance, to be appealing. However, it does not explain why some Y 1 failed the coding activity but managed to complete the programming design activity (see Figure 5) in which arrows also had to be drawn. Did some of the teachers help their pupils to complete the programming design activity during the post-tests, even when the representation of arrows had not been taught during the course of the program? This is a possibility.

If we set aside the idea that the teachers helped their pupils to complete the programming design activity during the pre-test and post-test phases, a second possibility emerges: the coding activity can be summed up as a simple translation where a tracing is converted into a set of instructions. On the other hand, the programming design activity requires imagination and mental representation. The Y 1 must indeed first construct the robot's trajectory to avoid the bird and reach the two flowers (see Figure 5). Then, in a second step, he/she must translate (code) the trajectory with instructions. The programming design activity may be facilitated by the five-year-old pupil's appropriation of meaning and meaning construction, something which the coding activity does not offer.

All these hypotheses would need to be confirmed by new experiments in future research work.



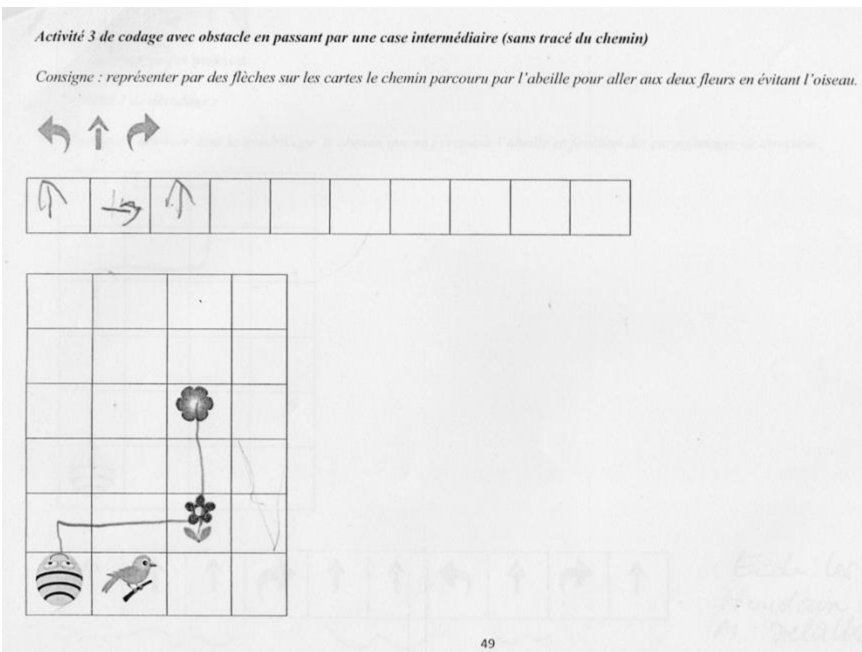


Figure 15. *These tests show arrows made by two different pupils*

6 Conclusion

The main objective of this study was to assess the reliability of the Blue Bot experiment through statistical means. This assessment made it possible to verify that the pre-tests and post-tests were conducted impartially. Then, it **also** verifies that the control group P (Placebo), which did not receive training in computer programming and robotics at school between the pre-test and post-test phases, presents lower performance rates than the other groups (see 2.2).

Since we were able to demonstrate **it** reliably, we then proceeded to analyze the data that emerged from the statistical processing. This allowed us to find the following results:

- With the exception of the BT and BRT groups, which we were not able to study, the B, R, T, BT and BR groups all revealed that, when compared to the control group P, there had been a significant progression between the pre-tests and the post-tests. These progressions were vastly superior to the P group; this is the reason why the preliminary activities and the various teaching modalities offered to the pupils had a significant effect. This means that teaching sequential programming to five-year-old children in their Year 1 is possible.
- In the case of the decoding and programming design activities, the Robot and Tablet (RT) modality presented the best performance rates. We must however note that the results of the RT bi-modality propose at the same time the worst score for low and regression (see 4.1.2.).
- Overall, the double modality hypothesis (see 4.2) seems to show that exposing Y 1 pupils to bi-modalities increases their learning, as revealed by the test performance scores for the decoding and programming design activities (see Figure 14). The time factor undoubtedly plays a preponderant role, but it is not the only factor involved for the coding activity since the use of the mono-modality or the bi-modality does not seem to change the situation in this particular case (see 4.2.3.).

While the Blue Bot project allowed us to report on results and to form hypotheses, there were also limits which would need to be considered in subsequent works:

- The BT group could not be studied due to a lack of data. Studying this group would help to confirm the hypothesis of the double-modality versus the mono-modality.
- Some groups, such as the BRT group, were underrepresented as they were only composed of 6 pupils. Therefore, we would need to find out whether 3 modalities could yield better results than the BR and RT modalities.
- The bi-modality hypothesis presented in discussions would need to be scientifically confirmed through specific experiments.
- The order in which the modalities were offered to the Year 1 pupils may have played a role in the results. It would also be necessary to check this fact.
- In the context of the Blue Bot project, another modality – paper – was presented to the pupils during the pre-tests and the post-tests. The tests may need to be adapted in order to take into account the modalities covered during the educational phases *ie*. the body, the robot and/or the tablet (see 1.2.).

- The 35 teachers involved in the Blue Bot project may have been influenced by the experiment.
- Overall, the experiment that we conducted in 2017 with the Blue Bot project was limited to a single geographical area, the Nord-Pas de Calais region in the North of France. Multi-centric studies would need to be conducted to confirm or disprove the results we obtained with a usable sample of 177 Y 1. To obtain more representative and significant numbers, studying a larger group of pupils that would cover France as a whole would be better.

Acknowledgements

For their advice, expertise, the sharing of their experience, support, translation and suggestions, thanks to: Anne Losq, Malika Brahmi, Joanne Frion, Emeline Visée, Gilles Brougère, Serge Tisseron, Sylvie Leleu-Merviel, Dorothée Hallier-Vanuxemm, Margarida Romero, Jean-François Condette, Alfonsino Cutillo, Romain Deledicq, Yoann Lebrun, Philippe Leclercq, David Detève, Angelino Mascaro, Patrick Pelayo and Gilles Petit, as well as all the partners associated with the Blue Bot research project: the Lille Dane, ID6, the academies of Lille and Dijon, the DeVisu Research Laboratory of the Polytechnic University of Hauts-de-France and INSPE Nord-de-France and the University of Laval.

References

Alvarez, J., Bellegarde, K., Flahaut, J-J. & Lafouge, T. *(2019). BLUE Bot Project experiment First statistical results of the performance of 5-year-old pupils on their ability to program, Proceedings of the 8th International Toy Research Association World Conference, International Toy Research Association (ITRA), hal-02090864, https://hal-univ-paris13.archives-ouvertes.fr/hal-02090864/document*

Augello, A., Daniela, L., Gentile, M., Ifenthaleri, D., & Pilato, G. (2020). *Robot-assisted learning and education. Frontiers in Robotics and AI*, 7, 147.

Bellegarde, K., Boyaval, J. & Alvarez, J. *(2019). S'INITIER à la robotique/informatique en classe de grande section de maternelle - une expérimentation autour de l'utilisation du robot Blue Bot comme jeux sérieux. ReSMICTE, 13(1), 51-72. https://pasithee.library.upatras.gr/review/article/view/3105*

Bers, M. U., Gonzales-Gonzales, C., & Armas-Torres, M. B. *(2019). Coding as a playground: Promoting positive learning experiences in childhood classrooms, Computers & Education, 138, 130-145.*

Cibois, P. *(1987). L'analyse factorielle, Collection « Que sais-je ? » Presses Universitaire de France.*

Depover, C., Karsenti, T. & Komis, V. *(2011). La recherche en éducation : étapes et approches. Chapter: La recherche évaluative, Publisher: Saint-Laurent, QC: ERPI Editors: Thierry Karsenti, Lorraine Savoie-Zajc, 213-228.*

Di Lieto, MC., Pecini, C., Castro, E., Inguaggiato, E., Cecchi, F., Dario, P., Cioni, G. & Sgandurra, G. *(2020). Empowering Executive Functions in 5- and 6-Year-Old Typically Developing Children Through Educational Robotics: An RCT Study, Frontiers in Psychology.*

Greff, JP. *(2004). Le corps d'abord !, Education enfantine, 1056, 62-63.*

Grible, J., Reimer, P., Rizo, A., Pauls, S., Cadwell, B., Macias, M., Spina, A., & Rosenbaun, L. (2020). Robot Block-based Coding in Preschool, In Gresalfi, M. and Horn, I. S. (Eds.), The Interdisciplinarity of the Learning Sciences, 14th International Conference of the Learning Sciences (ICLS) 2020, International Society of the Learning Sciences, Volume 4, 2229-2232.

Komis, V. & Misirli, A. *(2013). Etude des processus de construction d'algorithmes et de programmation par les petits enfants à l'aide de jouets programmables, In BARON G-L., BRUILLARD, E., & KOMIS V. Sciences et technologies de l'information et de la communication en milieu éducatif, edutice-00875628, 271-281.*

Leyzberg, D., Splauding, S., Toneva, M., Scassellati, B. *(2012). The physical presence of a robot tutor increases cognitive learning gains, in Proceedings of the Annual Meeting of the Cognitive Science Society (Sapporo) 34.*

Lowrie, T. & Logan, T. *(2006). Using spatial skills to interpret maps: Problem solving in realistic contexts, Australian Primary Mathematics Classroom, 12(4), 14–19.*

Nacher, V., Garcia-Sanjuan, F. & Martinez, J. J. *(2015). Game Technologies for Kindergarten Instruction: Experiences and Future Challenges, In CoSECivi, 58-67.*

Poletz, L., Encarnacao, P., Adams, K., & Cook, A. L. *(2010). Robot skills and cognitive performance of preschool children, Technology and Disability, 22(3), 117-126.*

Romero, M. & Loufiane (2016). *Vibot the Robot* , Publications du Québec.

Saporta, G. (2006). *Probabilités analyse des données et statistique*, Edition Technip, Paris.

Tremblay, N. *(2007). Formation initiale des enseignants, médiation pédagogique et approche philosophique, in TREMBLAY, N. (dir.), Des pratiques philosophiques en communauté de recherche en France et au Québec, Presse de l'Université de Laval (PUL).*