



HAL
open science

Could a Tangible Interface help a child to weigh his/her opinion on usability?

Julien Veytizou, David Bertolo, Charlotte Baraudon, Alexis Olry, Stéphanie Fleck

► To cite this version:

Julien Veytizou, David Bertolo, Charlotte Baraudon, Alexis Olry, Stéphanie Fleck. Could a Tangible Interface help a child to weigh his/her opinion on usability?. 30eme conférence francophone sur l'interaction homme-machine, Oct 2018, Brest, France. hal-01900032

HAL Id: hal-01900032

<https://hal.science/hal-01900032>

Submitted on 20 Oct 2018

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.

Could a Tangible Interface help a child to weigh his/her opinion on usability?

Est-ce qu'une Interface Tangible peut aider un enfant à « peser » son opinion sur l'utilisabilité d'un système ?

Julien Veytizou
Université de Lorraine - LCOMS
57070, Metz, France
julien.veytizou@univ-lorraine.fr

David Bertolo
Université de Lorraine - LCOMS
57070, Metz, France
david.bertolo@univ-lorraine.fr

Charlotte Baraudon
Canopé 57
57950, Montigny-Lès-Metz, France
charlotte.baraudon@reseau-canope.fr

Alexis Olry
Université de Lorraine - PERSEUS
57045, Metz Cedex 1, France
alexis.olry@univ-lorraine.fr

Stéphanie Fleck
Université de Lorraine - PERSEUS
57045, Metz Cedex 1, France
stephanie.fleck@univ-lorraine.fr

ABSTRACT

Research in the Human-Computer Interaction field needs to design interactions adapted for users and to rate opinion or personal perception of the interactions. Applied to the education field, one of the challenges of HCI is that it relies on the design of user-interfaces and interactions designed to meet pupils and teachers' requirements. To evaluate designed interactions, researchers currently use scale-based questionnaires like Likert's. For now, most of the time data collection is different between pupils and teachers. Indeed, young respondents have less experience than adults and could be more limited in their ability to express themselves. Several studies show that children tend to select extreme answers on these scales. To reduce this bias, we investigated the role of tangibility. We developed a prototypical device to help them to "weigh" their opinion. We conducted a preliminary study with children. First, our results show that our tangible interface seems to engage young respondents to meta-analyze system usability. Secondly our prototype seems to reduce extreme responses.

CCS CONCEPTS

Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

KEYWORDS

Tangible Interaction; Tangible User Interface (TUI); Usability testing; Evaluation design; Children; Bias limitation.

RÉSUMÉ

Les recherches en Interactions Homme-Machine nécessitent la conception d'interactions adaptées aux utilisateurs puis de les évaluer en recueillant entre autres l'opinion et la perception de ceux-ci. Appliquées au domaine de l'éducation, l'un des défis réside dans la conception d'interfaces utilisateur et d'interactions adaptées aux besoins des élèves et enseignants. Pour évaluer ces interactions, les chercheurs utilisent régulièrement des questionnaires basés sur des échelles comme celle de Likert. À l'heure actuelle, le recueil de données est souvent réalisé de façon différentes entre les élèves et les enseignants. En effet, les jeunes répondants ont moins d'expériences que les adultes et souvent rencontrent des limites pour s'exprimer. Plusieurs études montrent que les enfants ont tendance à choisir des réponses extrêmes sur ces échelles. Pour réduire ce biais, nous avons exploré une piste possible basée sur les interfaces tangibles. Nous avons développé un prototype pour aider les enfants à « peser » leur opinion. Nous avons ensuite conduit une étude préliminaire avec eux. D'une part nos résultats montrent que notre interface tangible semble engager les élèves dans une méta-analyse de l'utilisabilité du système évalué. D'autre part, notre prototype semble réduire le nombre de réponses extrêmes.

MOTS-CLEFS

Interaction Tangible, Interface Tangible (TUI), Test d'Utilisabilité, Conception de méthodes d'évaluation, Enfants, Limitation des biais

1 INTRODUCTION

This study is a part of the e-TAC research project (for Tangible and Augmented environment for Collaborative learning) funded by the French Ministry of National Education. This project aims

to enhance collaborative learning at school thanks to Augmented Reality (AR) and Tangible User Interfaces (TUIs) [15] (*n.b.* in accordance with Piaget's stages of development [45, 46], in this paper "children" refers to a specific age range: 7-11 years). Designing augmented and tangible environments, suitable and usable in the school context remains a stimulating Human-Computer Interaction (HCI) issue. One of the challenges lies in the design of user-interfaces and interactions tailored to the pupils' specificities, as well as those of their teacher. Indeed, as observed in the case of a pilot and a co-pilot in a cockpit [29, 30], children and adults are simultaneously users, *i.e.* co-users, of a same interface in a classroom. In contrast, and even if their actions are oriented towards learning, they do not perform the same activities, and have different expectations, needs and/or abilities. However, to be used in this context, interfaces must meet users' expectations and needs, whether the user is an adult or a child. Among other factors, and in relation to participative and user-centered design approaches, we therefore need to gather declarative information about how the pupils and their teachers perceive the usability, the appropriateness, and the attractiveness of augmented TUIs. However, some limitations could arise from scale-based questionnaires commonly used in usability testing of interactive systems when it involves children (see below). Beyond the limits associated with children and questionnaire based approaches, children are firm and could have a happier mind-set than adults [15]. Consequently, influenced by the halo effect that could be generated by TUI or AR, children, more than adults, could tend to select extreme answers on the scales. To reduce this bias, we investigated the role of tangibility.

This study addresses then the challenge of designing a novel evaluation solution by means of a new type of interactive and tangible Likert-scale called POP-IT (for *Personal Opinion Poll Interactive and Tangible*). This TUI aims to support the respondents by reminding them of the behavior of the assessed system instead of depending only on their emotion response, and the associated biases such as the halo or pitchfork effect. In this paper, after the presentation of the previous works that inspired us, we outline the principles of our prototype. Finally, we present and discuss the results of a pilot study involving 25 pupils (9-11 years-old). This study compared the differences between the answers to the Standard Usability Scale (SUS) mediated by an adult, obtained through a paper form *versus* POP-IT.

The main paper's contribution lies in the evaluation design choices supported by the promising results of the pilot study, opening a new field of possibilities for usability testing with children.

2 LIMITS OF SCALE-BASED USABILITY TESTING WITH CHILDREN

Some limits of user survey involving children may come from the survey method itself. Rating opinion or personal perception in usability testing currently uses scale-based questionnaires (*e.g.* Likert scales [41, 58], semantic differential scale [23]), which requires a graduated response to a series of statements. It is a common way to estimate attitudes, values, internal states and judgments [42]. Scales are commonly used in adult questionnaires. However, these scale-based assessments can have some limitations when used with younger participants [27, 40, 42].

From the age of 7 years old, children are able to participate in survey research, provided that the questionnaires are appropriately adapted [5, 7, 9, 51]. Among some response formats, the Likert scale is the one they prefer [36] but it is better to use 5-point scales [5, 22, 42]. Also, care should be taken to make scales clear and easy to interpret for children. To support children's understanding, many works recommend using a completely-labelled scale (*i.e.* every point has a label) with verbal labels or visual images (*e.g.* VAS for Visual Analogue Scales; SFL for Smiley Face Likert) rather than numerical labels [5, 7, 27]. However, young respondents have less experience than adults and be more limited in their ability to express themselves (*e.g.* less specific terminology, difficulty grasping an abstract concept, willingness to discuss with an adult [3, 28]). Even with alternative solutions such as oral, written and pictorial methods (see examples [40, 47]), because of these limitations children can have difficulty understanding, expressing their perceptions of the usability of an object or a system, and also being actively engaged in the rating task. Consequently, several studies have focused on finding a child-friendly scale format to measure children's judgments in a better way [22, 47, 59]. Unfortunately, most of the surveys developed for children are not validated and/or adapted to adult respondents. However, in the specific context of co-users involving adult and children, the scales must be similar for both populations to allow for comparison of answers.

Another limitation when evaluating TUIs or augmented environments, and more largely new interactive systems that provide a positive user experience, is the halo (or pitchfork) effect. The halo effect is a cognitive bias that causes a person to rate what he/she perceived (*e.g.* other persons, object, brand etc.) based on some ambiguous attributes [4, 43]. This effect could for example be generated by:

- the aesthetics and attractiveness of a person, an object, a brand etc. [11, 31, 39, 53]
- the novelty effect (sometimes call "waouh/wow effect") [35],
- the pleasure of participating in a session that is different from a "classic" activity, and even more if the interface is designed to provide an enjoyable experience [22].

The halo effect is yet an important lever to enhance learning by the improvement of attractiveness and by stimulating academic motivation [12]. Therefore, it should be specifically sought out in educative system design. However, this halo effect corresponds to an alteration of the users' judgments because the positive (or negative for a pitchfork effect) sensation generated could predominate over reasoning (see *e.g.* [43]). Indeed, if the user likes one aspect of the interface or of his/her user experience, he/she will have a positive predisposition toward everything about it. Then, this could make the respondents to tend to select extreme answers on the scales after the use of attractive interface for example [53, 55]. Unfortunately, this tendency is frequent when asking children about their experience [5]. As previously mentioned, children are firm and have a happier mind-set than adults. It is not therefore surprising that children only select positive ratings when involved in a pleasant experience [15].

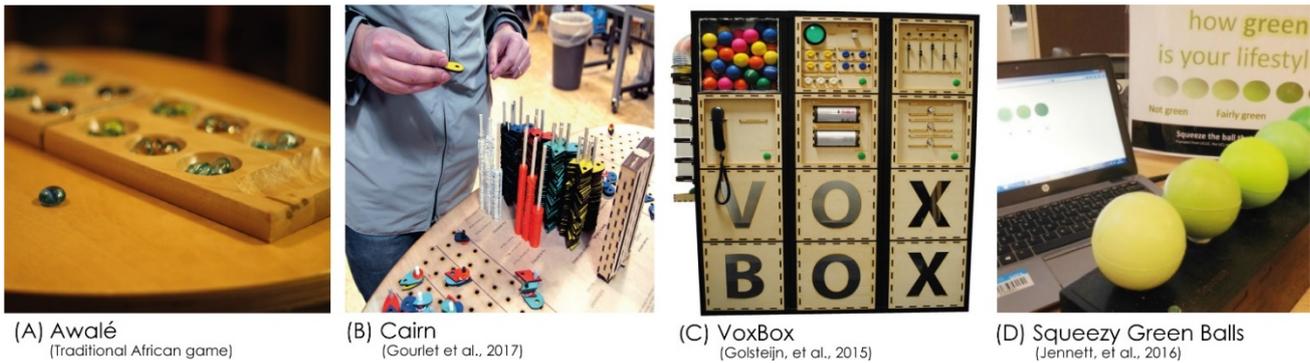


Figure 1: Examples of inspirational objects and tangible interfaces that require metacognition and the use of gestures to physicalize a personal choice. The Awalé/Oware (A) is a traditional game from the mancala family (see <https://en.wikipedia.org/wiki/Oware> for further information) and (B) Cairn [21], (C) VoxBox and its declination for children calls *SmallTalk* [17, 20] or Squeezy Green Balls [32] were designed to assess a personal experience: (B) in a Fablab, (C) after an entertainment, (D) of energy usage.

3 INSPIRATION AND DESIGN CHOICES

To overcome the issues exposed above, we propose to explore the potentialities of tangible interactions. In this approach, the concept of “*tangible*” has to be understood in its broad definition as an undeniable concept; real and not imaginary; able to be shown, touched, or experienced; and not only as an interface that provides a tangible representation of digital information and controls, but allows users to “grasp” data in their hands by physically manipulating these representations as defined by [Shaer and Hornecker \[52\]](#).

Even if children are able to answer a Likert-scale based survey, scale-based questionnaires (e.g. Likert-scales, semantic differential scales) most often remain unfamiliar to children. However, to be able to express his/her opinion (here about the usability of a system) in a balanced way, a user has to be aware of the intended goal of the survey. That is to say, the rating qualities of a system must be rendered tangible to the respondents. As demonstrated/showed/illustrated in [13], a considerable number of reviews highlight the promise of TUIs. TUI has many advantages to improve interaction accessibility, building on everyday skills and experiences [60]. They enable physical manipulation of artefacts that are relevant to the task, and they can make the contents to grasp less abstract, whether with children or adults.



Figure 2: Craft prototype of the POP-IT device.

In this context, a few research projects have been conducted to look at the benefits of TUI in evaluating tasks [17, 20, 25, 32]. These studies are mainly focused on tangible questionnaires that aim to catch users' attention. [Hofmann, et al. \[25\]](#) have developed *Bibox*, a tangible book rating and recommendation

system for a library. The main goal of *Bibox* was to motivate users' active participation. [Jennett, et al. \[32\]](#) have designed a playful installation call *Squeezy Green Balls* (see Figure 1.D) to encourage people to answer questions and engender feelings of concern about environmental awareness. [Golsteijn, et al. \[20\]](#) have developed a tangible form called *VoxBox* (see Figure 1.C) to gather public opinion and feedback during large public cultural events without being disruptive. They used 5-option spinners to represent a Likert-scale that made the user check out all the possibilities before answering.

All these studies were focused on adults. Based on *VoxBox*, [Gallacher, et al. \[17\]](#) have specifically designed another interface for children, called *SmallTalk*. This interface aims to assess children's experiences during cultural events.

Contrary to these studies, our project proposes to design only one system usable by both adults and children. Therefore, we designed a TUI calls POP-IT (for *Personal Opinion Poll, Interactive and Tangible*) to embody the tangible Likert-scale metaphors and the idea of weighing one's own opinion. This TUI resembles a kitchen scale (see Figure 2), which is an everyday object, easy to use and meaningful for both children and adults. It aims to afford the task to weigh. It's also intended to allow and stimulate respondents to be engage in gestural interaction. Finally, it is designed to support users in expressing their own perception of the usability of a system in a relevant manner.

Many arguments have led us to this solution. First, based on affordance theories [18, 44], providing meaningful contextual cues *via* an artefact could situate all the respondents in the expected action. But, as emphasized by [Hornecker \[26\]](#), “*the affordances of physical objects are potentially endless and users creatively select those that fit their understanding of the system, their aims and the situation.*” However, drawing on situated learning theories [38, 54], it is possible to create meaning from activities when connected to everyday life. This is possible even if the activity occurs in another environment (e.g. a classroom, a laboratory, a living lab). Therefore, a metaphor embodied by an everyday object can offer a direct link between what the object is for and the required task: expressing an opinion in a balanced way. Second, faced with the limitations when involving children in scale-based questionnaires, many previous works argued to adopt alternatives (see above). But, to our knowledge, the

transposition of gestures has rarely been used. However, gestures are naturally associated with verbal expression [1]. Moreover, using gesture when expressing something can have an impact on thinking due to the gesture's ability to reflect real-world actions [10, 19]. For example, it is the case with the various objects and interfaces that inspired us (see Figure 1). Thus, when someone uses Cairn [21] or plays with an Awalé, (see Figure 1.A & B) he/she must think carefully before and while making the gesture. Moreover, the requirements and consequences of gestural interaction, and the fact that they are public, compel respondents to meta-analyze. [37, 50, 56]. This has been observed for example when using tokens to support verbal participation in a classroom [6]. Indeed, interaction with the artefact, e.g. to place a seed in the Awalé or a token on Cairn, reflects player or user engagement; i.e. it reflects the person's active cognitive and psychological involvement in a task [2, 24, 48]. In our context, and as cited by Gallacher and Gallagher [16] "to do more than 'passively respond', a participant has to make a conscious decision to take part in a project or activity". Additionally, it is important to notice that engagement in classrooms decreases when the number of words, or the amount of text a student has to read increases [57]. The shape, the affordance, the implicit or explicit usage rules embedded in a tangible artefact could then provide stimulation and contextual cues to the user to enable and encourage him/her to interact with [33, 34]. Compared to paper format, a tangible interface could then provide meaningful elements to trigger the desire to interact and to cognitively engage respondents in achieving the rating tasks.

Therefore, designing an evaluation system based on gestural interaction with meaningful tangible artefacts that resemble an everyday object could:

- (i) support the respondents in becoming aware of how to rate the system (positively or not), without regard to its aesthetics or novelty;
- (ii) engage the respondents in making a meta-analysis and in expressing their opinion in a balanced way.

4 DESCRIPTION OF THE SYSTEM

The first version of POP-IT is a crafted prototype with foam board (see Figure 2). This technique allowed for a low-priced design. Inside, it includes an Arduino MEGA card to control all inputs and outputs electronic devices. The scale was built around a SEN0160 force sensor (1 kg) at Wheatstone Bridge. The sensor is coupled with a small breakout board which is based on the Hx711 circuit and a 24-bit precision analog-to-digital converter. The interaction with POP-IT is realized with the Seeed Studio 2,8" resistive touch TFT which integrates an SD card module. It is controlled by SPI communication. All users' results of scaled-based questionnaires are stocked in text files in the SD Card. Thus, the experimenters can recover all numerical data of their experimentation for later analysis and statistical processing. POP-IT also includes a lot of devices with Grove building block technology such as the Grove-Mega Shield to simplify connection and electrical connectors. With two push buttons, the experimenter can re-tare the scale and recalibrate the standard object used to calculate of the rate. This action can be executed at any time during the questionnaire sessions. In addition to electronic devices, POP-IT interface is made up of a glass

container in which people drop off wood marbles of the same weight.

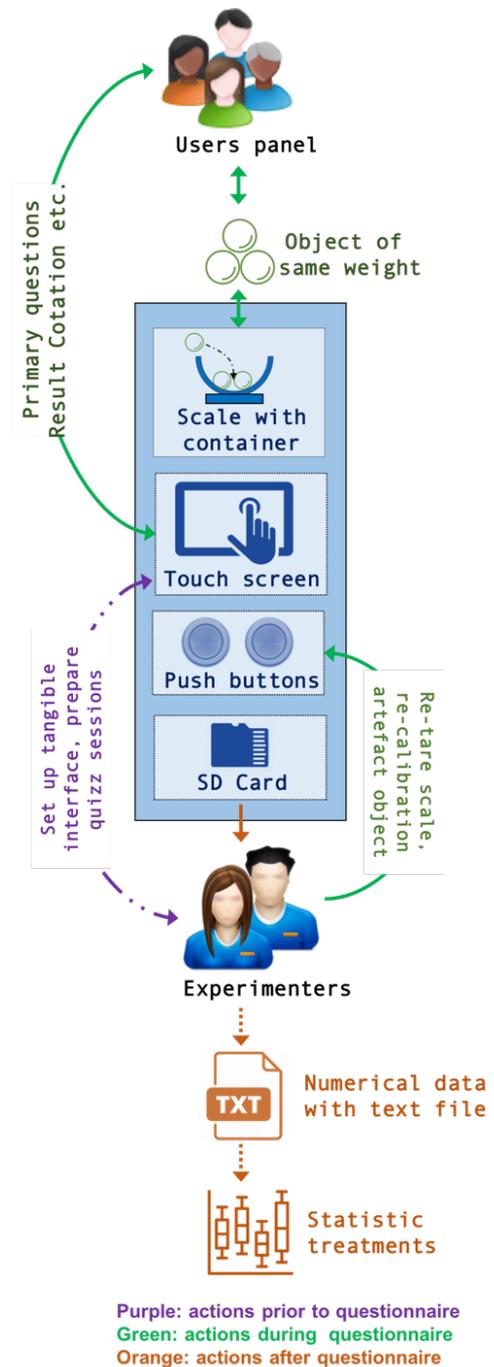


Figure 3: Description of the interactions typology from a user's point of view (i.e. children and experimenters).

At first, experimenters initialize the tangible interface and prepare the questionnaire sessions (see Figure 3). They leave the container on the scale. Once the system switches on, an automatic tare function is active. Then, experimenters drop a single wood marble to calibrate POP-IT and use the touch-screen interface to store the value. As soon as this step has been carried out, the pupils begin to answer the basic questions surrounding their gender, their age and their class from the touch-screen interface. This action allows them to familiarize themselves with

the system. Next, it is time for the pupils to give their opinion about the usability of a system by answering questions written on a paper questionnaire. They figure out the score of each question by depositing the number of objects corresponding to its result in the container. For example, if a question is to be rated from 1 to 5 and the pupils choose the rating 3, they place 3 objects in the container. The scale then measures the weight of their response. From the touch screen interface, the pupils can visualize and validate their results to proceed to the next question. Once the questionnaire is finalized, POP-IT offers each pupil a graphical presentation (*i.e.* chart bar) including the numerical value of their own score and of the average score of the cumulative results of all respondents.

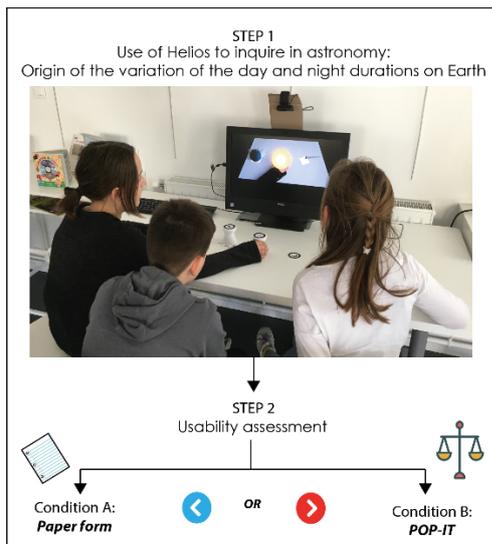


Figure 4: Session split in two steps. (1) Pupils solve an astronomy problem using Helios, (2) Pupils evaluate Helios's usability with a paper form or POP-IT.

5 PILOT USER-STUDY

We conducted a pilot user-study in real condition with pupils. It involved only children in order to determine if differences between the responses received via classic paper forms or via POP-IT could be already observable (or not) in a small population of children. The implementation of further researches in the direction of our hypotheses would be conditioned by its results. In this study we mainly verified if POP-IT:

- (i) engages young respondents to meta-analyze system usability;
- (ii) supports the children in becoming aware of how to rate the system, without regard to its aesthetics or novelty in order to reduce the halo effect.

5.1 Mean and methods

We deployed our current prototype during a half-day-long school outing dealing with the topic "First discovery of astronomy". The involved class comprised 25 pupils (9-11 years old, mean age: $M_{\text{age}} = 9.84$, $SD = 0.85$, 14 boys and 11 girls). The children had to solve the problem: "why is the duration of the day and night not always the same on Earth?" They used *Helios* to verify their own hypotheses. *Helios* is a tangible interface associated with marker-based Augmented Reality (AR). Its 3D visual environment is appealing and aesthetically pleasing in order to

stimulate the desire to learn. It currently generates strong positive emotions in children [14, 49].

The session was split in two steps (see Figure 4):

Step 1) Inquiring by using *Helios*;

Step 2) Evaluating *Helios*'s usability.

Implementing a user-study with pupils in real conditions of use, *i.e.* in a real educational context, lead to dealing with major constraints like schedule, respect of the curriculum, pedagogical support availability, or the pedagogical quality of the activities [12]. Therefore, it requires important instructional design work associated with the design of the user-study. Due to these constraints, we preferred a between-subjects design experiment for this pilot study.

The participants were divided in groups of 3 children (and one group of 4 in the paper condition), and switched workshops every 30 min. The evaluation of the usability of *Helios* was conducted through the French version of the *System Usability Scale* (SUS) [8], which uses a 5-points Likert-scale:

- Condition A. 13 pupils filled in the SUS on a paper form.
- Condition B. 12 pupils filled in the SUS with POP-IT;

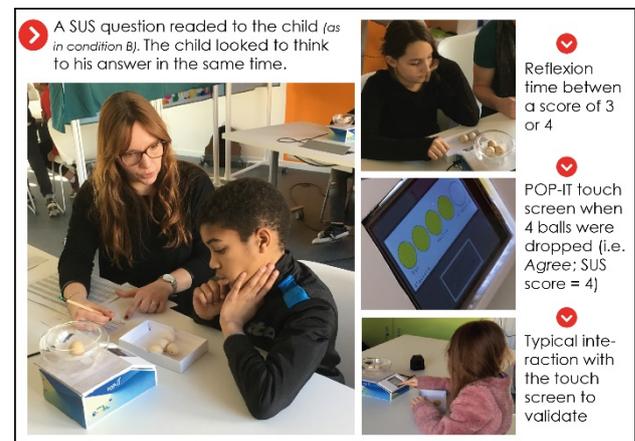


Figure 5: Evaluation process of the Helios's usability with SUS mediated by an adult by using POP-IT.

Because SUS sentences are complex to understand, according to [Read and MacFarlane \[47\]](#) recommendations, all the items of the SUS were read to the children in both conditions and reformulated by an adult to ensure that all the children had understood the sentence meaning (see Figure 5). In each condition, the pupil and adult-helper were paired randomly in order to avoid as much as possible a Pygmalion effect or other experimenter influences. All activities were video recorded in order to analyze pupils' behavior. One of the researchers analyzed the videos with the software "The Observer XT12". We used a behavioral analysis grid based on three main categories:

- Strategies for using the balls (taking or leaving the marbles in the container ...);
- Type of intervention of the experimenter (helping with regard to the interface, explaining the instructions, ...);
- Type of pupil task (manipulating the marbles, watching the score on the interface, doing nothing, listening to the instructions, ...)

We also defined and measured the "latency time" as the sum of the durations without manipulation between the end of the instructions and the pupil's answer.

5.2 Results and Discussion

5.2.1 Operability

The video recordings show few usability problems with POP-IT. Pupils were eager to manipulate the interface. The only usability difficulties for the POP-IT users came from the manipulation of the touch screen due to the size of the display. The information on the graphical interface was difficult to read and to handle for most of the children. In the next prototype, we will include Bluetooth technology to allow communication with a tablet. With a larger touch screen, we could then display a more usable environment. However, it doesn't prevent the children from responding.

5.2.2 Influence on children's behavior as respondents

The means (M) and medians of the filling duration of the SUS (i.e. latency time + filling in the paper form or placing the marbles + validation of the choice made only for the POP-IT users) are quite similar in both conditions (Paper form: $M_{\text{Paper}}=19,38$ min. $SD = 6,92$; POP-IT: $M_{\text{POP-IT}}=19,42$ min. $SD = 8,87$) without significant differences between the two conditions. Therefore, using POP-IT does not significantly increase the total duration of the usability test.

The video analyses indicated that all children according to their judgment dropped one or two marbles at a time into the glass container of POP-IT to express their opinion. Most of them left the marbles into the container after the validation of their answer. To answer the next question, 8 of the 12 children just added (or removed) the appropriate number of marbles. 3 children systematically left only one marble in the glass container before answering to the next question. 1 girl changed her strategy in function of the question.

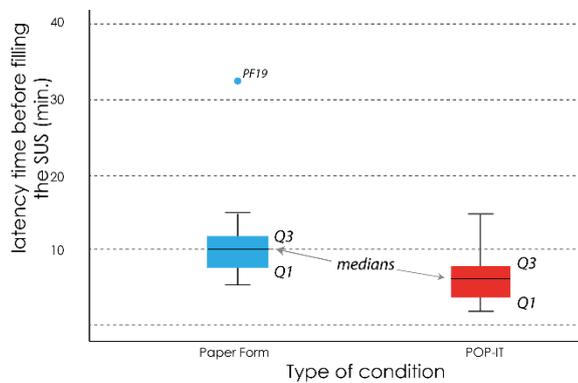


Figure 6: Comparison of children's latency time to express their opinion with a paper form (left) or POP-IT (right).

It is important to notice that 38% of the children in condition A (i.e. paper form) changed their answers after further explanations from an adult. Indeed, they frequently confused it with a written test of knowledge. They needed to be guided in order to understand how to fill in a Likert-scale in paper form. On the contrary, this need was not observed when using POP-IT. These behaviors are the first evidence that POP-IT is easy to use for children and more importantly could enhance children's awareness on how to rate usability using a Likert-scale. Moreover, total duration of responding is significantly correlated to the amount of time (latency) the children needed to determine their opinion (Pearson correlation coefficient "r" between total duration to answer the survey/child *versus* total duration of

latency/child: $r_{\text{Paper form}}=0,814$ - $p=0,001$; $r_{\text{POP-IT}}=0,610$ - $p=0,035$). However, except for one girl who needed a total of 32 min to make choices about how and what to answer to the 10 questions on her paper form (PF19 – see Figure 6), the children who filled in the SUS with POP-IT needed significantly less latency time to express their opinion on the usability of the interface than the one who used the paper form (see Figure 6 and T-test latency time *versus* condition types: $t=2,161$; $ddl=23$; $p=0,041$). In this specific condition of this pilot study, POP-IT then enabled respondents to meta-analyze system usability in a more efficient manner than by using a paper form.

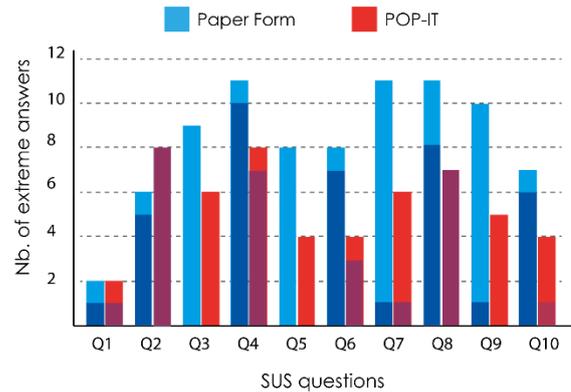


Figure 7: Number of extreme answers to each question of the SUS filled in a paper form (blue bars) or expressed by POP-IT (red bars). An extreme answer corresponds to a score of 1 (darker part of the bars) or of 5 on the 5-point Likert-scale.

5.2.3 Influence on Usability scores

The SUS scores results (Mean SUS scores: $M_{\text{Paper SUS}}=75.2$, $SD=14.41$ and $M_{\text{POP-IT SUS}}=68.3$, $SD=15.09$) confirm that *Helios* is perceived as usable by children. Moreover, no significant difference is noticed in the individual scores from SUS questionnaires in both paper form and POP-IT conditions (T-test $t=-1.16$, $ddl=23$; $p=NS$). Therefore, POP-IT doesn't significantly influence mean scores obtained. Consequently, the results obtained by using POP-IT could be compared to the ones obtained in paper form. As previously indicated, children's extreme answers increase with the halo effect. As illustrated by the Figure 7, extreme answers were frequent (Mean nb. of answers equal to 1 or 5 for the 10 questions of the paper form SUS: $M_{\text{Paper extreme answers}}= 8,30$ $SD=2,830$; $M_{\text{POP-IT extreme answers}}=5.40$, $SD=1.955$). It is important to notice that a T-test shows a significant difference between extreme answers in paper form and POP-IT conditions (T-test: $t=2.49$, $ddl=23$, $p=0.020$). Therefore, in the conditions of this study, POP-IT significantly reduces the children's tendency to select extreme answers on a 5-point Likert-scale.

CONCLUSION AND FUTURE WORKS

The purpose of this study was to identify if tangible interactions induced by a TUI as POP-IT could influence children's responses to questionnaires. Results provided evidence that this type of tangible interaction did not significantly influence the scores. This means that our TUI seems well suited for Likert-scale-based questionnaire. Our results also suggest that POP-IT enhance children's ability to "weigh" their responses. Indeed, extreme

answers were significantly reduced with the TUI than with the paper questionnaire.

Even if these are preliminary results, they support our hypotheses that tangible interactions could reduce the halo effect when children fill a Likert-scale based survey. This first study will allow for a more in-depth study.

The next steps will be then dedicated to:

- refining the POP-IT prototype to enlarge the interaction display. In future work, experimenters will be able to choose other types of scaled-based questionnaire and will access all setting of the TUI. To support children in their understanding of the sentences, the young users will listen to the questions on audio tracks recorded by the experimenters.
- validating the first results we obtained in this pilot study by expanding the panel. We will also duplicate it with adults to compare with the other co-users of interactive systems in a classroom to improve our prototype as an adult-child user friendly system. A within-subject will be designed and more parameters like the different objects usable as tangible artefacts will be also explored.

Beyond providing a new field of possibilities for usability testing with children, this study also open new opportunities to better include children with special needs such as dyslexic children, disabled peoples (e.g. with visual impairments) or newcomers in user studies.

ETHICS STATEMENT

This experiment was conducted after having obtained all legal agreements (data collection, photos, films and diffusion) from the headmistress, the teacher and the legal representatives of the pupils.

ACKNOWLEDGMENTS

The e-TAC project is financially supported by the French government in the e-FRAN call for projects within the framework of the Program of Investments for the Future (PIA) operated by Caisse des Dépôts et Consignation. Moreover, we would like to thank Estelle Laila, Julien Bonhomme and all the staff of Canopé 57 for their help. We also wish to give our thanks to all the children who took part along with their teacher.

REFERENCES

- [1] Martha W Alibali, Sotaro Kita, and Amanda J Young, 2000, Gesture and the process of speech production: We think, therefore we gesture, *Language and cognitive processes*, 15, 6, 593-613.
- [2] James J Appleton, Sandra L Christenson, Dongjin Kim, and Amy L Reschly, 2006, Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument, *Journal of School Psychology*, 44, 5, 427-445.
- [3] Lucy Atkinson, 2006, From play to knowledge: from visual to verbal?, *Anthropology Matters*, 8, 2
- [4] Neil E. Beckwith, and Donald R. Lehmann, 1975, The Importance of Halo Effects in Multi-Attribute Attitude Models, *Journal of Marketing Research*, 12, 3, 265-275.
- [5] Alice Bell, 2007, Designing and testing questionnaires for children, *Journal of Research in Nursing*, 12, 5, 461-469.
- [6] Kurt A Boniecki, and Stacy Moore, 2003, Breaking the silence: Using a token economy to reinforce classroom participation, *Teaching of Psychology*, 30, 3, 224-227.
- [7] Natacha Borgers, Edith De Leeuw, and Joop Hox, 2000, Children as respondents in survey research: Cognitive development and response quality 1, *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 66, 1, 60-75.
- [8] John Brooke, 1996, SUS-A quick and dirty usability scale, *Usability evaluation in industry*, 189, 194, 4-7.
- [9] Christine T Chambers, and Charlotte Johnston, 2002, Developmental differences in children's use of rating scales, *Journal of pediatric psychology*, 27, 1, 27-36.
- [10] Susan Wagner Cook, and Susan Goldin-Meadow, 2006, The role of gesture in learning: Do children use their hands to change their minds?, *Journal of cognition and development*, 7, 2, 211-232.
- [11] Alice H Eagly, Richard D Ashmore, Mona G Makhijani, and Laura C Longo, 1991, What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype, *Psychological bulletin*, 110, 1, 109.
- [12] Stéphanie Fleck, Charlotte Baraudon, Jérémy Frey, Thibault Lainé, and Martin Hachet, 2018, "Teegi's so Cute!": Assessing the Pedagogical Potential of an Interactive Tangible Interface for Schoolchildren. In *Proceedings of the ACM SIGCHI 17th International Conference on Interaction Design and Children IDC'18*, (Trondheim, Norway, 2018), 14 p.
- [13] Stéphanie Fleck, and Martin Hachet, 2016, Making tangible the intangible: Hybridization of the real and the virtual to enhance learning of abstract phenomena, *Frontiers in ICT*, 3, 30.
- [14] Stéphanie Fleck, Martin Hachet, and J. M. Christian Bastien, 2015, Marker-based augmented reality: Instructional-design to improve children interactions with astronomical concepts. In *Proceedings of the ACM SIGCHI 14th International Conference on Interaction Design and Children*, (Boston, MA USA, 2015), 21-28.
- [15] Stéphanie Fleck, Alexis Olry, David Bertolo, Christian Bastien, Robin Vivian, and Martin Hachet, 2017, Augmented and Tangible Environments: A Tool for Physicalization of Contents by Children in School Context? . In *Extended Abstract for Workshop: Pedagogy & Physicalization: Designing Learning Activities around Physical Data Representations at Designing Interactive Systems Conference (DIS '17)*, (Edinburgh, United Kingdom, 2017),
- [16] Lesley-Anne Gallacher, and Michael Gallagher, 2008, Methodological immaturity in childhood research? Thinking through participatory methods', *Childhood*, 15, 4, 499-516.
- [17] Sarah Gallacher, Connie Golsteijn, Yvonne Rogers, Licia Capra, and Sophie Eustace, 2016, Smalltalk: using tangible interactions to gather feedback from children. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 2016), 253-261.
- [18] William Gaver, 1991, Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1991), 79-84.
- [19] Susan Goldin-Meadow, and Sian Beilock, 2010, Action's influence on thought: The case of gesture, *Perspectives on psychological science : a journal of the Association for Psychological Science*, 5, 6, 664-674.
- [20] Connie Golsteijn, Sarah Gallacher, Lisa Koeman, Lorna Wall, Sami Andberg, Yvonne Rogers, and Licia Capra, 2015, VoxBox: A tangible machine that gathers opinions from the public at events. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, 2015), 201-208.
- [21] Pauline Gourlet, and Thierry Dass, 2017, Cairn: A Tangible Apparatus for Situated Data Collection, Visualization and Analysis. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, (Edinburgh, United Kingdom, 2017), 247-258.
- [22] Lynne Hall, Colette Hume, and Sarah Tazzyman, 2016, Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. In *Proceedings of the The 15th International Conference on Interaction Design and Children*, 2016), 311-321.
- [23] Mare Hassenzähl, Axel Platz, Michael Burmester, and Katrin Lehner, 2000, Hedonic and ergonomic quality aspects determine a software's appeal. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2000), 201-208.
- [24] E Tory Higgins, and Yaacov Trope, 1990, Activity engagement theory: Implications of multiply identifiable input for intrinsic motivation
- [25] Jacqueline Hofmann, Jens Müller, Bela Gipp, and Harald Reiterer, 2016, bibox: A Tangible Approach to Motivating Participation in Public Libraries, *Mensch und Computer 2016-Tagungsband*
- [26] Eva Hornecker, 2012, Beyond affordance: tangibles' hybrid nature. In *The TEI'12 Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, (Kingston, Ontario, Canada, 2012), 175-182.
- [27] Natacha Borgers, Joop J. Hox and Dirk Sikkel, 2003, Response quality in survey research with children and adolescents: the effect of labeled response options and vague quantifiers, *International Journal of Public Opinion Research*, 15, 1, 83-94.
- [28] Jean Hunleth, 2011, Beyond on or with: Questioning power dynamics and knowledge production in 'child-oriented' research methodology, *Childhood*, 18, 1, 81-93.
- [29] Edwin Hutchins, 1995, How a cockpit remembers its speeds, *Cognitive science*, 19, 3, 265-288.
- [30] Edwin Hutchins, 2001, Distributed cognition, International Encyclopedia of the Social and Behavioral Sciences. , Smelser, N., and Baltes, P. (Eds.), Elsevier Science.
- [31] <https://www.investopedia.com/terms/h/halo-effect.asp>, accessed april 2018.
- [32] Charlene Jennett, Ioanna Iacovides, Anna L Cox, Anastasia Vikhanova, Emily Weigold, Layla Mostaghimi, Geraint Jones, James Jenkins, Sarah Gallacher, and Yvonne Rogers, 2016, Squeezey Green Balls: Promoting

- Environmental Awareness through Playful Interactions. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 2016), 389-400.
- [33] Robert-Vincent Joule, and Jean-Lon Beauvois, 1989, Une théorie psychosociale: la théorie de l'engagement. Perspectives commerciales, *Recherche et Applications en Marketing (French Edition)*, 4, 1, 79-90.
- [34] John M Keller, 1997, Motivational design and multimedia: Beyond the novelty effect, *Strategic Human Resource Development Review*, 1, 1, 188-203.
- [35] H van Laerhoven, HJ van der Zaag-Loonen, and Bert HF Derkx, 2004, A comparison of Likert scale and visual analogue scales as response options in children's questionnaires, *Acta paediatrica*, 93, 6, 830-835.
- [36] Shirley Larkin, 2009, *Metacognition in young children*, Routledge.
- [37] Jean Lave, and Etienne Wenger, 1991, *Situated learning: Legitimate peripheral participation*, Cambridge university press.
- [38] Lance Leuthesser, Chiranjeev S Kohli, and Katrin R Harich, 1995, Brand equity: the halo effect measure, *European Journal of Marketing*, 29, 4, 57-66.
- [39] Panos Markopoulos, and Mathilde Bekker, 2003, On the assessment of usability testing methods for children, *Interacting with computers*, 15, 2, 227-243.
- [40] Ana Isabel Martins, Ana Filipa Rosa, Alexandra Queirós, Anabela Silva, and Nelson Pacheco Rocha, 2015, Definition and Validation of the ICF-Usability Scale, *Procedia Computer Science*, 67, 132-139.
- [41] David Mellor, and Kathleen A Moore, 2013, The use of Likert scales with children, *Journal of pediatric psychology*, 39, 3, 369-379.
- [42] Richard E Nisbett, and Timothy D Wilson, 1977, The halo effect: Evidence for unconscious alteration of judgments, *Journal of personality and social psychology*, 35, 4, 250.
- [43] Donald A Norman, 1999, Affordance, conventions, and design, *interactions*, 6, 3, 38-43.
- [44] Jean Piaget, 1964, *Six études de psychologie*, Denoël.
- [45] Jean Piaget, and Barbel Inhelder, 1969, *The psychology of the child*, Basic books.
- [46] Janet C Read, and Stuart MacFarlane, 2006, Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 conference on Interaction Design and Children*, 2006), 81-88.
- [47] Johnmarshall Reeve, Hyungshim Jang, Dan Carrell, Soohyun Jeon, and John Barch, 2004, Enhancing students' engagement by increasing teachers' autonomy support, *Motivation and emotion*, 28, 2, 147-169.
- [48] Benoit Roussel, and Stéphanie Fleck, 2015, "Moi, voilà ce que je voudrais que tu me fabriques !" (Lucie, 9 ans) : Design participatif pour l'utilisabilité de marqueurs tangibles en contexte scolaire. In *Proceedings of the 27ème ACM Conférence Francophone sur l'Interaction Homme-Machine-IHM 2015*, (Toulouse, 2015), a2 1-9.
- [49] Lise Saint-Pierre, 1994, La métacognition, qu'en est-il?, *Revue des sciences de l'éducation*, 20, 3, 529-545.
- [50] Jacqueline Scott, 1997, Children as respondents: Methods for improving data quality, *Survey measurement and process quality*, 331-350.
- [51] Orit Shaer, and Eva Hornecker, 2010, Tangible user interfaces: past, present, and future directions, *Foundations and Trends in Human-Computer Interaction*, 3, 1-2, 1-137.
- [52] Andreas Sonderegger, and Juergen Sauer, 2010, The influence of design aesthetics in usability testing: Effects on user performance and perceived usability, *Applied Ergonomics*, 41, 3, 403-410.
- [53] David Stein, 1998, Situated Learning in Adult Education *ERIC Digest*, ED418250, No. 195, 1-7.
- [54] Noam Tractinsky, Adim S. Katz, and D. Ikar, 2000, What is beautiful is usable, *Interacting with Computers*, 13, 2, 127-145.
- [55] Barbara Y. White, and John R. Frederiksen, 1998, Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students, *Cognition and Instruction*, 16, 1, 3-118.
- [56] Allan Wigfield, and John T Guthrie, 2000, Engagement and motivation in reading, *Handbook of reading research*, 3, 403-422.
- [57] Tao Yang, Jared Linder, and Davide Bolchini, 2012, DEEP: design-oriented evaluation of perceived usability, *International Journal of Human-Computer Interaction*, 28, 5, 308-346.
- [58] Yusrita Mohd Yusoff, Ian Ruthven, and Monica Landoni, 2011, The fun semantic differential scales. In *Proceedings of the 10th International Conference on Interaction Design and Children*, 2011), 221-224.
- [59] Bieke Zaman, Vero Vanden Abeele, Panos Markopoulos, and Paul Marshall, 2012, Editorial: the evolving field of tangible interaction for children: the challenge of empirical validation, *Personal and Ubiquitous Computing*, 16, 4, 367-378.