Universidad Politécnica de Madrid

ETS de Ingenieros Informáticos

**Usability testing guide for mobile applications focused on people with Down syndrome
(USATESTDOWN)**

Author: Ing. MSc. Doris Cáliz

Supervisor: Loïc Martínez

March, 2017

1	Content	
1.	Introduction	6
2	State of the matter.....	8
2.1	Usability and Usability testing	8
2.2	Down Syndrome.....	9
2.3	Bibliographic analysis of usability tests with people with DS.....	10
3	Preliminary experiments.....	15
3.1.1	Workshop Kiteracy Tool.....	16
3.1.2	Workshop Gesture Games	17
•	Task comparison	19
•	Completion time.....	21
➤	Success rate	22
4	Hypothesis and objectives	25
5	USATESTDOWN	27
5.1	Establish the tasks	29
5.2	Write the instructions	29
5.3	Define the test plan.....	30
5.4	Run the pilot test	32
5.5	Refine the test plan after analysing the results of the pilot tests.	33
5.6	Recruit participants	33
5.7	Run the test session	34
5.8	Analyse the collected information.....	35
5.9	Report results to the development team or management.	36
6	Evaluación	36
6.1	PHASE 1: Evaluating USATESTDOWN with ASSISTASK	38
6.1.1	Establish the tasks to be used in the usability tests.....	38
6.1.2	Write the instructions that participants will be given to perform the usability test	39
6.1.3	Define the test plan	39
6.1.4	Run the pilot test to analyze whether the process works to plan	41

6.1.5	Refine the test plan after analyzing the results of the pilot tests.....	41
6.1.6	Recruit participants	41
6.1.7	Run the test session.....	41
6.1.8	Analyze the collected objective (times, number of errors, etc.) and subjective (satisfaction questionnaires) data.....	41
6.1.9	Report results to the development team or management.	46
6.2	Evaluating USATESTDOWN through Money Handling tool.....	46
6.2.1	Evaluation of the Money-handling Training App – comparison of different age groups	48
6.3	Evaluating USATESTDOWN through PICCA2	49
6.4	Evaluating USATESTDOWN through EURECA	50
6.5	Conclusions from the evaluation of USATESTDOWN	50
6.5.1	Conclusions Phase 1.....	50
6.5.2	CONCLUSIONS PHASE 2	51
6.5.3	Conclusions Phase 3.....	52
6.5.4	Conclusions Phase 4.....	52
7	Conclusions and Results.....	53
7.1	Results	53
7.2	Conclusiones	54
8	Future Work.....	56
8.1	Publications	56
8.1.1	JCR.....	56
8.1.2	Conferences.....	56
9	Bibliografía.....	58
10	Anexos.....	61

Figure 1: Working Method Overview	7
Figure 2: Search refinement strategy flow diagram.....	11
Figure 3: Previous Experience	15
Figure 4: Diagram of the Tangible User Interface based on RFID sensors.....	16
Figure 5: Process to evaluate people with Down syndrome skills with gesture games app.....	18
Figure 6: Children taking part in the workshop with Gesture Games	19
Figure 7: Success rate by task.....	19
Figure 8: Mean completion time in milliseconds by task and age group. Error bars depict standard error.	21
Figure 9: Mean completion time in milliseconds by task and gender (F: Female, M: Male).....	21
Figure 10: Success rate by task and age group.	22
Figure 11: Success rate by task and gender (F: Female, M: Male)	23
Figure 12: Users grouped by the number of successfully performed tasks.	24
Figure 13: Process of User Centered Design.....	27
Figure 14: Process of Usability Testing as defined in USATESTDOWN	28
Figure 15: Working Method Overview	37
Figure 16: Applying USATESTDOWN	38
Figure 17: The test plan process	40
Figure 18: Success Rate per Document.....	43
Figure 19: Total time needed to process each document, expressed in seconds	43
Figure 20: Questions and acceptance percentages of SUS participants	44
Figure 21: Acceptance rate percentages of the questions for the tutors	46
Figure 22: Workshop process for testing MHT App with USATESTDOWN	46
Figure 23: Mobile devices used for testing usability depending on the device size; Tablet with 22.2" and mobile phone with 4.7" running Android 6.	47
Figure 24: Money Handling with Adults with Down Syndrome	48
Figure 25: PICAA 2 with people with Down Syndrome.....	49
Figure 27: EUREKA with people with Down Syndrome	50

Propuesta de una guía de evaluación de usabilidad para aplicaciones móviles centrada en personas con síndrome de Down.

(USATESTDOWN).

Las pruebas de usabilidad de las aplicaciones móviles que involucran a las personas con síndrome de Down es una cuestión que no ha sido exhaustivamente investigada y no hay una sola propuesta que asuma todos los temas que deberían ser tomados en cuenta [1]. Este estudio tiene como objetivo proponer una guía práctica "USATESTDOWN" para evaluar la usabilidad de las aplicaciones móviles orientadas a usuarios de síndrome de Down y sus principales limitaciones. El estudio comienza con un análisis de las metodologías y herramientas existentes para evaluar la usabilidad e integra los conceptos relacionados con los métodos de inspección e investigación en la propuesta, la misma incluye las opiniones de expertos y usuarios representativos; sus limitaciones, la aplicabilidad durante el proceso de desarrollo y la accesibilidad. Esta guía se basa en la revisión de la literatura y la experiencia del autor en varios talleres donde más de 108 personas con síndrome de Down utilizaron dispositivos móviles con aplicaciones específicas para evaluar diferentes parámetros y destrezas. Una vez creada la guía se procedió a evaluarla en un proceso iterativo con diferentes herramientas y centros especiales para personas con Síndrome de Down. Como resultado de este proceso tenemos una guía robusta y adaptada especialmente a la realidad y necesidades de las personas con DS.

Usability testing guide for mobile applications focused on people with Down syndrome (USATESTDOWN).

The usability testing of mobile applications involving persons with Down syndrome is an issue that has not been comprehensively investigated and there is no single proposal that takes on board all the issues that could be taken into account [1]. This study aims to propose a practical guide "USATESTDOWN" to evaluate the usability of mobile applications focusing on Down syndrome users and their primary limitations. The study starts with an analysis of existing methodologies and tools to evaluate usability and integrates concepts related to inspection and inquiry methods into a proposal. The proposal includes the opinions of experts and representative users; their limitations, the applicability during the development process and the accessibility. This guide is based on the literature review and the author's experience in several workshops where 108 persons with Down syndrome used mobile devices with specific software applications to evaluate different skills and parameters. After we created the guide we evaluated this with an iterative process with different tools and special Centre to people with Down syndrome. As a result of this process we have a robust guide adapted especially to the reality and needs of people with DS.

1. Introduction

Today mobile technology use is increasing, in fact the use of mobile phones is already overcoming the use of traditional computer. Additionally various organizations from government, the private sector and academia are becoming more interested on the opportunities opened by technology to assist people with Down Syndrome to live better lives in their daily lives. [1]

Increasing attention has recently been drawn in the Human-Computer Interaction (HCI) community towards the design and development of accessible computer applications for individuals with developmental or cognitive impairments [2]. In this way it is important that these technologies can be used by all people, including people with disabilities. To this end, accessibility guidelines have been defined (such as WCAG, EN 301 549) which define requirements that must be fulfilled in order for an TICs (The technologies of the information and communication) product to be accessible [3].

There is a group of people with disabilities who are not well represented in those guidelines, which are people with cognitive disabilities. And more specifically, the group of people with Down syndrome, who will be the focus of the Thesis [4]. There is interesting research on supporting people with different conditioning profiles but still much more to be explored, in particular there is no much reported on the specific case of supporting people with Down Syndrome (DS) [1].

Most of the software, games, and Web sites that persons with Down syndrome interact with are designed without consideration of their special needs, making the applications less effective or completely inaccessible[2]. Given the special characteristics of people with Down syndrome it is essential that mobile technologies have a very high degree of usability.

In order to carry out this process, the specific needs of this population group must be taken into account. Research has found that children and young adults with Down syndrome can use computers effectively (including the mouse, keyboard, and screen), often without any assistive technologies or modifications [5]. The skills that individuals with Down syndrome need to use computers include fine motor skills, visual-motor skills, visual memory skills, letter recognition skills, reading and literacy skills, and, depending on how instructions are given to them, auditory and visual processing skills [6].

To achieve this high degree of usability it is necessary to evaluate the usability, through the usability test, involving participants with Down syndrome[7]. There are a number of recent examples of studies involving technology usage by individuals with Down syndrome. Although case study design, survey, ethnographic observation, and experimental methodology have been used to study computer use in the population, usability testing has not yet been employed as a method for investigation of computer use in people with Down syndrome as we will show on the State of the Question.[6]

Therefore, the general objective of this Doctoral Thesis is to define and evaluate this methodological proposal, taking on count this important points explained previously. This guide has been called USATESTDOWN.

- **Working method overview**

In general, the working method has four main phases, as shown in Fig. 1. The process is iterative.

- **Theoretical analysis.** A state of the art on usability testing involving persons with Down syndrome. The purpose is to identify good usability testing practices and possible guidelines for this process when participants are people with this cognitive disability. These practices and guidelines should account for their specific impairments.
- **Experimental analysis.** There have been made workshops on usability testing with 122 persons, 69 children and 53 adults with Down syndrome to determinate the skills, behave and how they interact with mobile devices. Participants were recruited from several daycare Centre's in Spain (Madrid and Valencia). We evaluated
- **The guideline "USATESTDOWN":** This phase consists on the preparation of a guideline to perform usability testing involving persons with Down syndrome. The contents of the guideline, called "USATESTDOWN" are based on the results of phases 1 and 2. The development of the guideline will be iterative. Observational evaluation has been chosen as the method to be used in the usability testing.

- **Evaluation of “USATESTDOWN”:** The USATESTDOWN guideline will be evaluated with a set of experiments involving persons with Down syndrome in usability testing of mobile applications. The results of the evaluation will be used to improve the guideline.

Figure 1: Working Method Overview

The Phases 1 and 2 are described as the State of the matter .The Phases three is USATESTDOWN guide. The phase 4 is USATESTDOWN evaluation.

2 State of the matter

2.1 Usability and Usability testing

- Usability

- Definition: NIELSEN define the usability is a quality attribute that assesses how easy user interfaces are to use. The word "usability" also refers to methods for improving ease-of-use during the design process [8].
- Components : Usability is defined by 5 quality components:
 - Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
 - Efficiency: Once users have learned the design, how quickly can they perform tasks?
 - Memorability: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
 - Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
 - Satisfaction: How pleasant is it to use the design?[8].

However the International Organization for Standardizations (ISO) bases usability on three main attributes: effectiveness, efficiency and satisfaction. Systems with good usability are easy to learn, efficient, not prone to errors and generate user satisfaction [9].

- Usability Evaluation Methods

There are several methods for evaluating how usable a product or system is: heuristic or guideline evaluation, usability testing and follow-up studies of installed systems [10].

- Inspection or Heuristic: According to Hornbaek in [11], there are several approaches to usability evaluation , we describe the most importants:
 - Heuristics: It involves usability expert evaluator. Assessment is done based on a checklist or prepared questionnaire
 - Observation: The evaluator will assess how users interact with the computer and the software in questions. The evaluator will note the interactions which can represent success, failure, happiness and dissatisfaction displayed by users
 - Interview: Meetings between the evaluator and users were arranged, and interviews were conducted to obtain information relating to the needs, methods that can be used, opinion as well as responses towards the courseware.
 - Cognitive Walkthrough: A task-oriented method. Testing is done based on a checklist which has been prepared. Emphasizes on cognitive issues, such as “learnability “, by analyzing the mental processes required by the user
 - Pluralistic Walkthrough: An extension of cognitive walkthrough. Users, developers, and usability experts go through the interface and discuss on the elements concern.
 - Think Aloud: It involves having an end user continuously thinking out loud while using the software. By verbalizing their thoughts, the users allow the evaluator to understand how they view the software, which makes it easier to identify the end users’ misconceptions.
- Follow-up studies of installed systems: We can use the following options

- Questionnaires (Interviews, surveys)
 - Field Studies (Direct Observation, Audio and Video Recording)
 - Focus Groups
 - Real Use Record
 - Customer Attention Services (Online)
- Usability Testing: The most common method for evaluating how usable a product or system is usability testing, which involves testing prototypes with real users [4]. Participating users are given a set number of tasks that they have to perform using a prototype or a full system. Data on the effectiveness, efficiency and satisfaction of users are collected during testing. A user-centred design process is applied to build products and systems with a satisfactory level of usability [9]. As part of this process, planning, context of use analysis, interactive system design and evaluation tasks are carried out iteratively. A key step is usability evaluation.
- Definition: Usability testing (UT) is an evaluation method where the performance of typical users is measured as they carry out real, pre-defined tasks using the target application.
 - The aim : UT is to test the usability of the system, not the users' ability. UT is expensive, requiring sophisticated usability laboratory equipped with monitoring cameras and equipment[12]. Usability testing involves testing prototypes with real users. Participating users are set a number of tasks that they have to perform using a prototype or a full system. Data on the effectiveness, efficiency and satisfaction of users are collected during testing [13].
 - Steps : The usability process is divided into the following steps:
 1. Recruit participants after determining the population group of interest and the required number of participants.
 2. Establish the tasks that are to be used in the usability tests.
 3. Write the instructions that participants will be given to perform the usability test.
 4. Define the test plan, which is a protocol stating activities like welcome, pre-test interview, observed task performance by user, satisfaction questionnaire, personal interview to gather qualitative information, etc.
 5. Run the pilot test to analyse whether the process works to plan.
 6. Refine the test plan after analyzing the results of the pilot tests.
 7. Run the test session.
 8. Analyse the collected objective (times, number of errors, etc.) and subjective (satisfaction questionnaires) data.
 9. Report results to the development team or management.

2.2 Down Syndrome

Down syndrome (DS) is a genetic disorder with a worldwide incidence close to one in every 700 births but the risk varies with the mother's age. Persons with DS have impaired cognitive processing, language learning and physical abilities, as well as different personal and social characteristics [14]

The characteristics (phenotype) for persons with trisomy 21 may include: delayed development of expressive language , receptive language that is superior to expressive language , relative strength in vocabulary and pragmatics, greater difficulties with morphology and syntax , difficulties with short term memory and recall with specific difficulty with memory for verbal information, strengths in visual processing , difficulties with complex conversational skills, and reduced speech intelligibility.

There are a range of functional abilities in individuals with Down syndrome, related to the extent of impairment in the various sensory and motor channels, memory, cognition and communication skills, they have also problems in the sensory areas of hearing , vision, and tactile (touch) [7]. In the area of motor skills, both fine motor (e.g. cutting with scissors) and gross motor (walking) skills are delayed.. Visual processing and visual memory are superior to auditory processing and auditory memory [4].

Jinjuan Feng [15] presents a research with experiences with persons between 4 and 21 years old with Down syndrome. It was made a survey asking how persons with Down syndrome make use of computers. In specific questions about input device, the most cited were keyboard (85.6%) and mouse (93.2%). Other input devices cited were: touch screen (12.3%), joystick (7.5%), touchpad (5.5%), trackball (4.9%), speech recognition (3.4%), stylus (2.3%) and keyguard (0.4%). One of the conclusions of the work was persons with Down syndrome have much difficult to use mouse and specially keyboard because they have fingers shorter than usual. In addition, a better understanding of usage patterns may lead, in the short term, to the development of more structured experiments and research studies, and in the long term, more usable mobile applications.

2.3 Bibliographic analysis of usability tests with people with DS.

We applied a review and document analysis methodology with two protocols: one for searching for sources of information and the other for inspecting the sources of information [16]. Table 1 shows the search protocol and Table 2 illustrates the document analysis protocol.

Table 1: Information source search protocol

Information source search protocol		
Language:	English	
Period:	2008 to 2014	
Term	Individual	Usability, evaluation, down syndrome, cognitive disabilities, hci, human computer interaction
	Combinations	Search 1: USABILITY EVALUATION DOWN SYNDROME Search 2: COGNITIVE DISABILITIES USABILITY
Information resources	WEB OF SCIENCE UAM, INGENIO UAM, COPUS UAM, GOOGLE ACADEMICO, MICROSOFT ACADEMIC SEARCH, ERIC, REFSEEK, SCIENCE RESEARCH, WORLD WIDE SCIENCE, SCIELO CERN, SCIENCE DIRECT, SCIENCE, ACM AND SPRINGER	
Search strategies	Two searches were run with combinations of different keywords: <ul style="list-style-type: none"> Search 1: “usability evaluation” and “down syndrome” Search 2: “cognitive disabilities” and “usability” The results were successively refined considering: <ol style="list-style-type: none"> Year of publication: from 2008 to 2014 Relation of publications to technologies and computing Relation of usability to computer systems usability (Human-Computer Interaction – HCI). 	

We analysed the usability testing previous contributions for mobile applications focused on persons with Down syndrome. Generally, the usability process is divided into the following steps: 1. recruit participants, 2. establish the tasks, 3. write the instructions, 4. define the test plan, 5. run the pilot test, 6. refine the test plan, 7. run the test session, 8. analyse the collected objective, and 9. report results

The literature review process (Figure 2) was composed of two main searches and combinations: one used the terms “usability evaluation” and “down syndrome” and the other employed the terms “cognitive disabilities” and “usability”. The preliminary list of papers (621 + 415) was first pruned based on date of publication and the relevance of paper titles. This returned 58+57 papers. The list was further pruned based

on the relevance of the content of the abstracts. The result was a list of 98 papers (43 + 55). These papers were read and analysed, and 11 papers were found to be of relevance to the topic of usability testing for persons with DS.

Figure 2: Search refinement strategy flow diagram

The literature review process has consisted in two searches, one with terms “usability evaluation down syndrome” and the other with the terms “cognitive disabilities usability”. The initial list of references was pruned in a first stage based on the relevance of their titles. Then a second pruning was made based on the relevance of the content of the abstracts. The result was a list of 98 papers. These papers have been read and analysed, then we had 11 articles.

These 11 papers were thoroughly analysed and sorted by priority (high, medium or low) depending on their contributions to the steps of the usability testing process. The result was a list of five high-priority papers.

We showed the Information source inspection protocol in table 2.

Table 2: Information source inspection protocol

Information source inspection protocol	
Inspection rules:	<p>The order of inspection is as follows:</p> <ol style="list-style-type: none"> 1. Inspection of title 2. Inspection of abstract 3. If the information is relevant to the research topic, the content is inspected.
Exclusion criteria:	<ol style="list-style-type: none"> 1. Duplicate information 2. Information unrelated to the research topic 3. Outdated information.
Inclusion criteria:	<ol style="list-style-type: none"> 1. Information relevant and related to the research topic

We applied a new filter giving a priority and an important level to the contribution research taking in count the approach of the investigation to the actual research. See Table 3. Finally we obtained result 5 papers have been useful to extract information about usability testing with participants.

Table 3: Giving a priority to the most important researches

DOCUMENT	PRI ORI TY	SUMMARY
A method to evaluate disabled user interaction: a case study with Down syndrome children [17]. 2013.	High	This study designed by [17] evaluated four children aged between 6 and 12 years with DS and analyses the development of the coding scheme based on the detailed video analysis method (DEVAN) to observe the interaction of the children with DS. Also applies IQ evaluation and use JECRIPE tool. The test plan is to deliver the application to the children, observe and film. No pilot test was run. Finally, the workshop was held and the results for each child evaluated on average for 45 minutes for all process were analysed.
A Usability Evaluation of Workplace-Related Tasks on a Multi-Touch Tablet Computer by Adults with Down Syndrome [6]. 2012.	High	Two pilot sessions are run: administer demographic questionnaire to participants and validate participant recruitment criteria. Participants were asked to perform five different categories of tasks on an iPad (social networking, electronic mail, scheduling / planning, price comparison and basic text input / note taking). No formal data collection or methodology was applied. Use patterns were observed. They were then used to write a list of tasks and develop a methodology. Participants were reevaluated during the second session, and this information was used to rewrite the list of tasks.
Designing Usability Evaluation Methodology Framework of Augmented Reality Basic Reading Courseware (AR BACA SindD) for Down Syndrome Learner [18]. 2011.	High	This paper proposes a usability evaluation framework for an augmented reality framework for learners with DS. To do this, three to five expert interface design and learning content evaluators were recruited. They analysed 10 adults with DS to evaluate how proficient they were at using multi-touch tablets for job-related tasks. The evaluation was divided into two phases: an acceptance testing phase including formative assessment and a usability phase including either a formative phase with an iterative development cycle or a summative phase where testing is conducted with a large number of users. The goal was to identify strengths and weaknesses [18].
The complementary role of two evaluation methods in the usability and accessibility evaluation of a non-standard system [19]. 2010.	High	[19] worked with five usability and accessibility experts and six learners to evaluate a literacy system in Africa. It was evaluated using the heuristic method and a usability field study. First a pilot study was run to gain an idea of how the applications work. The pilot study activities were: run the evaluation and draft a report of the compiled evaluation for submission to the immediate evaluator.
Usability Evaluation of Multimedia Courseware (MEL-SindD) [20]. 2009.	High	This paper discusses the usability assessment of the courseware, the methods used for the evaluation, as well as suitable approaches that can be deployed to evaluate the courseware effectiveness for disabled children. The evaluation was divided into three phases: PHASE 1. Identify user needs, PHASE 2. Evaluate usability with the participation of 11 students with DS, and PHASE 3. Send the data collected by the researcher to the specialist teachers and parents of the recruited children with DS.
Usability of the SAFEWAY2SCHOOL system in children with cognitive disabilities]. [21]	Low	Fourteen children with DS and a control group of 23 children without disabilities participated in the study conducted by (Falkmer et al., 2014) which involved evaluating a system for improving safe school transport for children.
Validating WCAG versions 1.0 and 2.0 through usability testing with dis-abled users [3]. 2012.	Low	This paper reports a study that empirically validated the usefulness of using WCAG as a heuristic for website accessibility.
Usability remote evaluation: METBA system [22]. 2012.	Low	This paper reports a solution (METBA) for managing the information related to the evaluation of human behavioural observation . The system is used to register and manage the information derived from remote usability evaluation and complements the methodology commonly used in this research area.
Computer Usage by Children with Down Syndrome: Challenges and Future Research [2]. 2010	Low	This paper reports the text responses collected in the survey and is intended as a step towards understanding the difficulties experienced by children with DS when using computers.
A multi-method, user-centered evaluation of accessibility for	Low	The Study have assessed the accessibility of web site from federal e-government. The conclusion is that web sites should be accessible to persons with disabilities.

persons with disabilities [23]. 2009.		
Computer Usage by Young Individuals with Down Syndrome: An Exploratory Study [4]. 2008.	Low	This paper discusses the results of an online survey that investigates how children and young adults with DS use computers and computer-related devices.

We found finally 5 articles related with our topic after a Literature Review research. We used the definition of the main steps of usability testing to analyze the contributions of each author on each usability testing step. We took the authors findings in each point [7], [8], [9], [10], [11]. We had the results in table 1. We can see there are several empty spaces, they mean there are not contribution in those specific field.

The literature review process described in Section 3 focused on identifying papers that report a usability test with people with Down syndrome and on retrieving the key information that they provide on each of these nine steps. The Table:4 , show the detailed contribution of each author in each phase of the usability process.

Table 4 : Part of the analysis of the research on usability testing for people with DS

1. Recruit participants	From the analysis of the research with regard to the recruitment of participants, we find that [17] take four children aged from 6 to 12 years with DS, [19] use five usability experts and six learners, [18] use from three to five interface design and learning content experts, and [24] work with two paediatricians, primary school teachers and 11 children with DS. This illustrates the importance of working with on average 10 paediatricians, interface and learning content evaluators and people with DS.
2. Establish tasks	[17] Holds a 30-minute training session, takes 20-minute videos per child and uses the DEVAN method to work directly with children with DS. On the other hand, [19] evaluate a literacy portal in Africa using the following tasks: submission of evaluation criteria, submission of document stating procedure to be followed, submission of document on interfaces and applications for evaluation, signature of anonymity and confidentiality forms. In the research by [19], the experts identify critical usability problems in the early stages of the development cycle and divide the evaluation into two phases: acceptance testing and usability. [20] Divide the tasks used in the evaluation into several phases: PHASE 1. Identify user needs, iteratively engage students in testing, and collect data from teachers and parents of students with DS, PHASE 2. Conduct the usability evaluation, and PHASE 3. Collect the data from specialist teachers and parents and hold the scheduled interviews. The activities specified by [6] are validate the criteria for recruiting participants, like computer experience.
3. Write instructions	[20] Describe the instructions for identifying the needs of users, which are collect data, interview students' paediatrician and primary school teachers, interact socially with students; identify the learning needs. Understand the problems through conversations with parents; interview specialists, teachers and parents as informers on the background of students and the research.
5. Pilot testing	[19] Conduct a pilot test aimed at understanding how applications work. [6] Believe formal data collection to be important for the pilot test. This should be followed by a second session during which they suggest modifying the list of tasks, adding a warm-up task, giving tips on how to move forward and encouraging thinking aloud.
7. Testing	[20] Collect the data iteratively from people with DS in Phase 1. Another aim is identify the suitability of the teaching material for the learning problems that students are set. [19] Describe the testing steps: execute evaluation, write report, submit report to immediate evaluator, okay report, and compile evaluation reports.

After the exhaustive analysis we wrote the contributions of each paper Table 5 sets out the information regarding which papers provide key information for each of the steps.

Table 5: Previous contributions of usability testing

Paper	1. Participants	2. Tasks	3. Instructions	4 Define Test Plan	5. Pilot testing	6. Testing	7. Testing	8. Analyse	9. Report
[17]2013	X	X							
[19]2010	X	X			X.		X		
[18]2011	X	X							

[6]2012		x			x				
[20]2009	x	x		x			x		

Note that there are contributions regarding five of the nine usability testing steps: recruit participants (1), establish tasks (2), Define Test (4) Plan pilot testing (5) and testing (7). Table 5 contains the key contributions regarding each of the steps.

Briefly, the retrieved information is as follows. As regards the instructions on tasks, there is very little information. Additionally, the test plan that can be enacted for the population group of interest is not clearly defined. Even though pilot testing greatly improves the second round of testing, pilot tests are seldom used, and the papers fail to establish the format or steps to be taken. As regards testing, they only describe the activities performed without any specific specifications for participants with DS. Therefore, we can conclude that the different papers contain no recommendations as regards the addressed research topic. Table 5 details the activities to be performed to achieve the specific goal of each piece of research but not a general-purpose method proposed by the authors that is applicable across the board.

The analysis reveals that usability has been well researched. As regards usability evaluation, there are many proposals and methodologies. However, we have not found any significant efforts considering mobile applications and persons with DS. On this ground, there is a patent need to state guidelines on all the steps to be taken to test the usability of applications for mobile devices for persons with DS

The main findings of this literature review is that mobile applications usability testing including persons with Down syndrome is an issue that has not be comprehensively investigated. While there is some related research, this is incomplete, and there is no single proposal that takes on board all the issues that could be taken into account. Consequently, we propose to develop guidelines on the usability testing process involving participants with Down syndrome.

We developed a paper with this analysis and it was published on the ICAIT 2016 Conference[25].

3 Preliminary experiments

We performed several workshops in different Special Dow Syndrome Centre in Spain as we show on **Figure 3** . We involved a total number of 122 participants, 69 children and 53 adults with Down syndrome to determinate the skills, behave and how they interact with mobile devices.

- Tool :
 - Kiteracy Tool: Kiteracy is an educational kit designed to improve the literacy process of persons with Down syndrome by enabling higher levels of interaction.[26] This prototype consists on a modular smart kit to support literacy acquisition for children with intellectual disabilities[27] The smart kit could generate a natural language experience where children are taught the language they will need in their everyday lives. Educators could configure a number of challenges to give children opportunity to explore and discover.[28].
 - Gesture Games: Is an application with the basics activities on a touch screen superficies. This tool take the time and the Success rate. The interaction framework for the experiment was implemented in Java using JMonkeyEngine SDK v.3.0beta. The devices used for deployment and the experiment was a Samsung Galaxy Note 10.1 tablet with Android 3.2 both with capacitive multi-touch screens [29].
- Down syndrome Special Centre:
 - Asyndown Centre The aim is to facilitate and promote all the necessary means to achieve the integral development of persons with Down syndrome and to promote their full family, educational, work and social integration as active members of society. [30]
 - Maria Corredentora: Special education Centre to persons with Down syndrome. [31]
 - Apadema : Association for persons with intellectual disabilities and Down Syndrome [32]
 - Prodis: This is an institution without profit objectives , they try to improve the school, social and labor integration of persons and young persons with intellectual disabilities and Down Syndrome [33]

PREVIOUS EXPERIENCES				
TOOL	PLACE	PARTICIPANTS	AGE	TOTAL
Kiteracy	Asindown Centre Valencia - Spain	9 M 5 F	5 – 10 years old	69 Children
Gesture Games	Ma. Corredentora Centre . Madrid - Spain	20 M 20 F	5 – 10 years old	
	Asindown Centre Valencia - Spain	11 M 4 F	5 – 10 years old	
Gesture Games	Apadema Centre Madrid - Spain	15 F	27 -54 years old	53 Adults
	Prodis Centre Madrid - Spain	16 M 22 F	19 – 40 years old	
56 F + 66 M = 122 Persons				

Figure 3: Previous Experience

3.1.1 Workshop Kiteracy Tool

Literacy acquisition is essential for the rest of the life for a child who has intellectual disabilities. In this context, technology can support specific strategies that will help children to learn to read. This workshop presents a Tangible User Interface (TUI) using Radio Frequency Identification (RFID) sensors to support literacy for children with Down syndrome. The research focuses in the use of physical objects to give children an interactive and physical sensory experience to develop their literacy skills. The design of the TUI framework is based on the integration of RFID sensors with tags embedded in 3D printed objects and low cost toys. The work presents the experience of using some materials covering the tags and the evaluation of the TUI framework in a special education institution. We evaluated 14 children with down syndrome [27].

Structured experiments were performed to assess the perceived usefulness and usability of our prototype in real scenarios. We carried out the evaluation in a special education institution with five teachers and twelve children with Down syndrome. To perform the evaluation, teachers were trained on how to use our system and they were who incorporated it into the learning process that they normally follow. Each child participated in two, 15 minute sessions.

In order to evaluate the design of the prototypes we carried out we conducted a thinking-aloud test [34] followed by an interview about the user experience with five special education teachers. The goal was to understand their experience while teaching children with Down syndrome. This includes the method, type of educational resources and teaching strategies they used to teach children how to read. For this purpose, we explained how each prototype works and asked them to verbalize their thoughts, simulating a real scenario. The test helped us to identify some problems related to the size of pictograms, the type and color of the words, the type of voice and the time for automatic pronunciation.

In the first session, the system was set up with the following components: one Laptop (Processing unit), two RFID readers, a set of 24 recognizable objects, and a tablet or interactive board. The teacher carried out some reading activities related to the method and manipulated the different versions of the prototypes separately (Cards, GUIs and TUIs). The teacher selects one card and shows to the child pronouncing the associated word. The child has to repeat the pronunciation and identify the concept. In the GUI and TUI versions the teacher also interact in the same way. In the second session, we designed an activity oriented to develop self-learning. During this session the child interacted with the system without the presence of the teacher. All of these sessions were recorded in video for subsequent analysis. **Figure 4**

Figure 4: Diagram of the Tangible User Interface based on RFID sensors

This section describes the design and development strategy we followed to build our proposed TUI system. The development processes was supported by five special education teachers. We used an Iterative User-

Interface Design to develop prototypes and to consider some usability metrics, such as, easy to learn, efficient to use, easy to remember and pleasant to use [27].

- **Conclusions**

After the validation process, all of the teachers agreed that the system is useful in enhancing the first stages of the literacy process. One of the teachers mentioned that the system could help not only children with intellectual disabilities, but also children with visual disabilities when touching and recognizing the objects. The teacher believes that the adaptation of novel technology is necessary to the integration of children with disabilities at the classroom.

Teachers also thought that tangible interfaces emphasize the connection between the body and cognition, facilitating thinking through physical actions. Tangible media is a complement to literacy method because it gives digital information and physical shape to the studied words.

Is important to mention that most of the participants in the evaluation process pointed the fact that the system could be used to help the literacy process for any child, including children without disabilities.

A key component of learning strategies to develop reading skills is the interaction. In this sense we analyzed two learning strategies: teacher-child (the teacher works with one child) and self-learning (the child works alone without the presence of the teacher). In each strategy we carried out a random tests during each session: Card (paperboard cards), GUI (digital cards displayed on tablet or interactive board) and TUI (Tangible RFID objects). Table 1 shows the results of the lever of participation between teachers and children.

The feedback of the test enabled us to improve the system with the following features:

(a) The color of the physical object are the same of the pictogram in the GUI. (b) The color of the words in cards and GUI are red. (c) The size of the pictogram and the words are big as possible. (d) Each word is preceded by the article. (e) The GUI has an option to change the font, color and size of the words. (f) The voice of the system is female.

We got a JCR at Sensors Journal (IF = 2.245) with the topic Creating TUIs Using RFID Sensors—A Case Study Based on the Literacy Process of Children with Down Syndrome [27].

3.1.2 Workshop Gesture Games

The overall goal of our study was to investigate the suitability of a basic set of multi-touch gestures for DS persons. Using the GQM (Goal Question Metric) [34] template [35], it could be defined as follows: analyse a set of multi-touch gestures for the purpose of evaluating their suitability from the viewpoint of usability in multi-touch technologies in the context of DS persons.

DS persons of both genders aged between 5 and 54 years old were considered. In order to find out to what extent different levels of motor cognitive development have an impact on their ability to interact with multi-touch devices and according to the usage of age as main factor in other related literature consisting of studies that analyze motor and cognitive skills development in persons with DS [36] [37].

The performance of the several groups was analyzed to find any differences between them. Since development differences have been reported between male and female persons in early development (e.g. higher mastery of motor skills in girls or higher performance of boys in spatial tasks [38]. We were also interested in exploring possible gender differences in performance; hence, the participants were balanced in terms of gender, so that Gender and Age group were the two main independent variables considered. The Completion time and the Success rate were the two measured dependent variables for each task (tap, double tap, long press, drag, scale up, scale down, one-finger rotation).

The following research questions were defined:

- RQ1: Is the degree of success independent of the task?
- RQ2: Is the completion time independent of the task?
- RQ3: Is the success rate of task k affected by gender?
- RQ4: Is the success rate of task k affected by age group?
- RQ5: Is the completion time of task k affected by gender?
- RQ6: Is the completion time of task k affected by age group?

The first two questions were related to the homogeneity of the success rates and completion times of the different tasks and the other four were defined for each task performed (type of gesture).

3.1.2.1 Participants

In order to answer the research questions listed above, fifty-five DS children aged between 5 and 10 years took part in the experiment (Mean (M) = 6.67, Standard Deviation (SD) = 1.50). The genders of the participants were balanced, with 30 males and 25 females. The children were divided in two age groups: 5 to 7 years (12 males and 12 females), 8 to 10 years (18 males and 13 females). The children involved in the experiment were from special centers for DS children. These centers apply personalized and individualized educational methodologies to help DS children develop new cognitive and motor skills. All children had previous experience using parental smartphones.

3.1.2.2 Equipment

The interaction framework for the experiment was implemented in Java using JMonkeyEngine SDK v.3.0beta. The device used for deployment and the experiment was a Szenio 10.1 tablet with Android 4.2 with a capacitive multi-touch screen.

3.1.2.3 Process

For each task, the children were given a 5-minute learning phase with an instructor. The experimental platform then asked them to perform the task without any assistance (Figure 5). They had to perform three repetitions of each gesture under specific conditions. When the gesture was completed successfully, the platform gave a positive audiovisual feedback. If the instructor saw that the participant did not carry out the task in a given time, it was marked as undone and the child went on to the next one.

Figure 5: Process to evaluate people with Down syndrome skills with gesture games app.

For each interaction, the system recorded the start time (seconds needed to go into action after the visual stimulus was shown), completion time, success (performed correctly or incorrectly), and the number of contacts with the surface (in order to know in an unsuccessful action whether the user had made any attempt to interact). A qualitative analysis was also carried out from the notes taken by an external observer during the experimental sessions. We show childrens participating on the workshop on the Figure 6.

Figure 6: Children taking part in the workshop with Gesture Games .

3.1.2.4 Results

In this section, the results of the experimental tests are presented according to each of the analysed independent variables.

- Task comparison

The three tests carried out by each participant for each task were also combined in order to depict the success as a percentage. If a participant performed successfully either zero or one tests in a specific task, he (she) was considered incapable of performing it, whereas if they successfully performed two or three tests in a specific task, they were considered capable of doing it, as they actually showed their ability to consistently reproduce the gesture several times. According to this codification the success rate for each task is shown in **Figure 7**, in which it can be seen that, except for long press, all the tasks were performed with nearly a 100% of success rate. See

Figure 7: Success rate by task

A pair-wise task comparison was conducted to test whether degree of success was independent of task. A Pearson's chi-square test of independence was used. Table 6 shows the results of the statistical analysis. Each cell contains the significance obtained from the analysis of each pair of tasks (*, $p < 0.05$; **, $p < 0.001$).

Table 6: Task comparison by success with Pearson's chi-square test of independence χ^2
(DoF=1, N = 55)

Success	Tap	Double Tap	Long Press	Drag	Scale Up	Scale Down	1-Finger Rotation
Tap		0.315	0.000 **	1.000	0.315	0.315	0.079
Double Tap			0.000 **	0.315	1.000	1.000	0.308
Long Pressed				0.000 **	0.000 **	0.000 **	0.001 *
Drag					0.315	0.315	0.079
Scale Up						1.000	0.308
Scale Down							0.308

As can be seen in **Figure 7**, the results in Table 6 showed that not all the gestures had the same success rates. According to the statistical tests, the first gesture category (tap, double tap, drag, scale up, scale down and one-finger rotation) was defined in which the children achieved similar performance in terms of a success rate close to 100 per cent. The second group was made up of the long press gesture in which the children had some issues and achieved a lower success rate of around 70 per cent. The variance in completion time of each task is presented in Table 7 and was also analyzed.

Table 7: Variance of completion time (in seconds) by task

		Variance
Task	Tap	3.15
	Double Tap	6.57
	Long Press	15.66
	Drag	3.81
	Scale Up	33.70
	Scale Down	15.96
	One-Finger Rotation	10.77

Table 8 shows Levene's test for the homogeneity of variance in the completion time when comparing the tasks. Each cell contains the significance of a combination of two tasks. In this case $\chi^2(\text{DoF}(\text{Task X, Task Y}) = \text{DoF}(\text{Task X}) + \text{DoF}(\text{Task Y}) - 2$ (see Table 9 for the DoF values of each task). Given the large number of comparisons (a family of $m=21$ hypotheses) we applied a Bonferroni correction that establishes statistical significance at $p < 0.05/21 = 0.00238$.

Table 8. Results of Levene's test

Success	Tap	Double Tap	Long Press	Drag	Scale Up	Scale Down	1-Finger Rotation
Tap		0.102	0.000 **	0.946	0.016 *	0.000 **	0.000 **
Double Tap			0.005 *	0.135	0.102	0.001 *	0.009 *
Long Pressed				0.000 **	0.917	0.678	0.479
Drag					0.018 *	0.000 **	0.000 **
Scale Up						0.704	0.751
Scale Down							0.211

In the results of the analysis shown in Table 8, two groups of gestures can be defined regarding the variance in the children's completion time. The first group includes Tap, Double Tap and Drag, which are the gestures

with the lowest level of variance, i.e. the results are homogeneous in terms of completion time for these gestures.

The Long Press, Scale Up, Scale Down and One-Finger Rotation gestures compose a second group which represents the tasks with a higher level of variance. This implies that the completion times are more disperse and there are some differences between the subjects when performing the task.

- Completion time

In order to aggregate the data per participant and task, the three trials carried out by each one per gesture were combined. The average of each subject's successful tasks was used to obtain the average completion time value per task and user. If the test was not performed successfully it was not included in the completion time analysis, resulting in different statistical degrees of freedom for each task. The mean completion time for each task by age group and gender is given in **¡Error! No se encuentra el origen de la referencia.** and **¡Error! No se encuentra el origen de la referencia.** (see **¡Error! No se encuentra el origen de la referencia.**) and graphically in Figure 8 and Figure 9.

Figure 8: Mean completion time in milliseconds by task and age group. Error bars depict standard error.

Figure 9: Mean completion time in milliseconds by task and gender (F: Female, M: Male).

A two-way between-subject ANOVA with the independent variables gender and age group and dependent variable completion time was applied. The statistical analysis (Table 9) demonstrated that the completion

time was not significantly influenced by age group in any of the tasks. The analysis also showed that the completion time of all the tasks was not significantly influenced by gender.

Table 9: F-Statistics of the completion time analysis

Task	DoF	Gender		Age group		Gender*Age group	
		F	p-value	F	p-value	F	p-value
Tap	(1.55)	1.861	0.179	0.043	0.836	0.119	0.732
Double Tap	(1.55)	0.545	0.464	0.081	0.777	5.478	0.023
Long Press	(1.48)	0.455	0.504	0.520	0.475	0.980	0.328
Drag	(1.55)	1.834	0.182	0.620	0.435	2.097	0.154
Scale Up	(1.55)	1.009	0.320	0.781	0.381	0.006	0.940
Scale Down	(1.54)	0.229	0.635	1.040	0.313	0.916	0.343
One-Finger Rotation	(1.55)	1.406	0.241	1.099	0.300	2.131	0.151

➤ Success rate

In order to obtain one value per user and task, the success data was combined as detailed in the Task comparison section. The resulting success rate of each task is shown in **Figure 10** and **Figure 11**

Figure 10: Success rate by task and age group.

Figure 11: Success rate by task and gender (F: Female, M: Male)

Pearson's chi-square tests were conducted on each gesture to determine the independence of success from two qualitative factors (gender and age group). The tests showed that degree of success and age group are independent in the Tap, Double Tap, Drag, Scale Up, Scale Down and One-Finger Rotation tasks (see Table 10). However, they also showed that in the Long Press task there is no empirical evidence to confirm that degree of success and age group are independent, since the older age group performed the task with a significantly higher success rate than the younger group.

Table 10: Statistics of Pearson's chi-square test

Task	DoF	Gender		Age group		DoF	Gender*Age group	
		χ^2	p-value	χ^2	p-value		χ^2	p-value
Tap	1	0.000	1.000	0.000	1.000	3	0.000	1.000
Double Tap	1	1.222	0.269	1.316	0.251	3	3.650	0.302
Long Press	1	0.182	0.670	4.441	0.035	3	4.801	0.187
Drag	1	0.000	1.000	0.000	1.000	3	0.000	1.000
Scale Up	1	0.849	0.357	0.789	0.375	3	2.094	0.553
Scale Down	1	0.849	0.357	1.316	0.251	3	3.650	0.302
One-Finger Rotation	1	0.576	0.448	0.684	0.408	3	1.550	0.671

Regarding the gender factor, there is no empirical evidence to verify that degree of success is not independent of gender (see Table 10). Hence, degree of success does not have a significant relationship with gender in any of the evaluated tasks. The analysis also shows that there is no empirical evidence to confirm that degree of success is not independent of the joint gender*age group.

Figure 12: Users grouped by the number of successfully performed tasks.

In **Figure 12** we can see a histogram of the number of users able to perform a given number of tasks successfully (ranging from 0 to 7). On one hand, only one user failed in more than one task; and in fact failed in three of the tasks (Long Press, One-Finger Rotation and Scale Down). On the other hand, we can see that 98% of the users were able to perform 6 or 7 of the evaluated multi-touch gestures. The children in this group failed especially in the Long Press task. In this case, we observed that they had an entry precision problem that prevented them from keeping their finger in a fixed position at the start of the interaction. Instead, they performed a drag around the entry point, where the finger then remained pressed. This makes the system misinterpret the initial contact, because the start of a drag gesture prevents it from identifying a long-pressed gesture, no matter how long the finger is kept pressed.

3.1.2.5 Observational findings

In addition to the automatic data logging that was performed to measure completion times and degree of success, an external observer gathered valuable information regarding the behavior of persons during the experiments. These observations revealed different precision issues that will now be described.

The main issue that the external observer reported occurred in the Long Press task in which persons have to tap an element with one finger and hold it for a given amount of time to succeed. In this case, persons hold their finger on the target but sometimes they perform little undesired movements with their finger which are read by the system as drag events. This precision issue prevents persons to complete the task successfully. When this situation occurs, persons feel upset, frustrated and make visible their disenchantment with the game by saying that it is not working properly. Right away, they lift their finger from the tablet and retry to succeed in the task by pressing very hard on the target. This makes the task more difficult since the harder they press the more likely to perform undesired drag events.

Another recurrent situation was the cognitive complexity associated with the process of counting in the double tap interaction. Several persons tapped repeatedly on the screen 4 or 5 times even when the target had disappeared instead of only tapping the 2 taps required to succeed in the task. According to educators there are two possible reasons to explain this fact; the first one is that persons were unable to count the number of required events whereas others could be able to count but had motor-inhibition problems that prevent them to stop after two taps.

In the gestures which require two contact points to be performed such as Scale up and Scale down the participants who tried to pass the task by using both hands sometimes had some issues since they tend to rest their hand on the surface and this prevented the system to correctly perceive the scale gesture. So, some persons experiencing this issue changed their strategy to succeed in the task and used two fingers of the same hand (the index and thumb) pointing out that this strategy was more comfortable for them.

Finally, two issues were observed in the rotation task. The first one relative to the performance of the gesture; although the task was designed to be performed only with one contact, two users tried to accomplish the task with two hands preventing them to succeed. The second issue reported was relative to the efficiency when performing the gesture; most of the persons rotate the target in clockwise direction despite the fact that sometimes the shortest way to pass the task was to rotate in the anticlockwise direction. This could be because they were right handed and using clockwise rotations they were not occluding the target element with their hands.

These issues will be discussed below in the context of the interaction aids or design guidelines that application designers should take into account if these touch interactions are included in future applications.

3.1.2.6 Quantitative findings

The overall goal of our study was to investigate the suitability of a basic set of multi-touch gestures for DS persons. The workshop evaluated the abilities of these persons with down syndrome (aged from 5 to 55 years) when performing a basic set of multi-touch gestures (tap, double tap, long press, drag, scale up & down, rotation) in tablet devices. The results show that regardless of their more limited motor skills, DS persons are able to perform most of the evaluated multi-touch gestures with success rates close to 100 per cent and that this technology could be fully exploited for developing applications targeted specifically at this type of user.

Only the long press gesture, in which the persons scored a 70% success rate, presented some issues. In this case, the persons (mainly the youngest) did not have the required motor skills to press the target and hold their finger on it. An assisted strategy to filter out undesired spurious events at the beginning of the interaction could therefore be included in order to adapt this gesture to the person actual motor skills.

The results obtained from this work suggest that DS persons are able to perform a basic set of multi-touch gestures when no precision is required in the termination phase of the gestures. However, some trends of future work arise and need to be addressed by researchers in the near future. For example, it remains to be explored whether they are able to use these gestures when high levels of accuracy are required in the termination of the gesture. Future work may be carried out to ascertain whether DS persons's motor skills fit with applications which require high levels of precision when performing the gestures. In the present study the gestures were evaluated in isolation and more research is needed to determine whether these persons are able to perform these multi-touch gestures in sequences.

Finally we can say

- Multi-touch technology is used in most emergent devices and is widely accepted
- DS persons are able to perform a basic set of multi-touch gestures
- The long-press gesture is not possible by DS persons without assistive strategies
- Gender differences are not significant in interaction success rates

To this section we wrote a paper to the International Journal of Human-Computer Studies Journal (IF = 1.268) with the topic "Examining the Usability of Touch Screen Gestures for Children with Down Syndrome" it was approved and we made the corrections the reviewers requested . It was reviewed two times.

4 Hypothesis and objectives

The usability testing of mobile applications involving persons with Down syndrome is an issue that has not been comprehensively investigated and there is no single proposal that takes on board all the issues that could be taken into account [17]. However research towards supporting independent living for persons with intellectual disabilities is not so extensive. Within this target group, user needs vary widely, and there is potentially a very broad range of e-Accessibility and usability challenges that need to be addressed. These

sensory and motor issues would need to be taken into consideration when investigating computer usage and accessibility needs in individuals with Down syndrome [7]. As an analysis result a usability testing methodology suitable for participants including persons with DS needs to be well designed [39].

We propose a usability guide adapted to the special skills of the down syndrome persons taking as base the 9 steps of the the Usability Testing Process.

To each step of of the the Usability Testing Process the guide propose specifics Guidelines inside this point to take into account the special needs of participants with Down syndrome.

Some of these guidelines include adapted versions of questionnaires' submitted to participants, they were designed with the Down syndrome participants Tutors. The Tutors were expert's people working with persons with Down syndrome. Questionnaires' such as SUS USATESTDOWN was adapted only with 3 levels of answers and also we modified the sentences into an easy way to understand to the participants

The adequacy of the USATESTDOWN guide will be evaluated, applying it to usability assessments of mobile TICs systems, and collecting the assessment of participants with Down syndrome and their educators or tutors.

Because persons with DS have special characteristics, they need high levels of usability of the products they use. For a design with good usability approach design user centered, involving users need several phases, including "usability testing". To persons with DS is important that the process of usability testing suits according to the needs of these persons. To solve this problem we propose USATESTDOWN.

5 USATESTDOWN

We describe the guideline to perform usability testing involving persons with Down syndrome. USATESTDOWN is based on the Process of User-Centered Design. Standard:ISO 9241 -210: 2010 [40]as shown in **Figure 13**.

Figure 13: Process of User Centered Design.

The contents of the guideline, called “USATESTDOWN” are based on the results of chapter 1 and 2 of this document. The development of the guideline will be iterative. Observational evaluation has been chosen as the method to be used in the usability testing. This study aims to propose a practical guide “USATESTDOWN” to measure and evaluate the usability of mobile applications focusing on Down syndrome users and their primary limitations.

USATESTDOWN is a guide to help usability tests of mobile applications focused on users with Down syndrome. Applying the usability testing guide USATESTDOWN, the evaluators can easily use the usability test recommendations to evaluate on a real case an application developed to work with persons with Down syndrome. We took the 9 steps described before but now we explain what is especial on persons with Down Syndrome when they will use a application on a mobile device. The flow of the process was adjusted to account for the reality of the persons with Down syndrome. We showed in **Figure 14** , the USATESTDOWN Guide process.

Figure 14: Process of Usability Testing as defined in USATESTDOWN

After the USATESTDOWN figure scheme we have 9 tables from number II until X with the following information:

- **Step:** The step name based on the Usability Test process
- **Definition:** According with the authors [18], [19], [20] what we should expect in this point.
- **Bibliography Research:**
 - **Bibliography Suggestion:** A collection about what the authors propose in this step [7] [8] [9] [10] [11].
 - **Real Experience:** The real results obtained after to apply the author guideline. It is the information collected with several workshops, it means the experience that we collected with this experiments. The workshops were explained on chapter 2.
- **Usatestdown:**
 - **UsatestDown Guideline :** Contribution of the proposal Guide. This part contain the experience before and after to create UsaTest down because we tested the proposal guide with several workshops and we improve every time making finally a complete guide.
 - **Got From Experience :** The experience we had after perform the workshops and the reason why we propose the guideline. It is the justification or reason to propose the recommendation. This
- **Documents:** Documents to support the step adapted specially to persons with DS.

5.1 Establish the tasks

Table 11: Usatestdown:

USATESTDOWN		
Step:		
Definition	This step consists of defining the tasks that the participants will complete in the usability tests. These tasks will be identified in pre-development phases to identify which of them will form a part of the evaluation that will include tasks that appear in certain usability specifications, as well as other representative tasks.	
Bibliography Research	BIBLIOGRAPHY SUGGESTION	REAL EXPERIENCE
	<ul style="list-style-type: none"> [17] Holds a 30-minute training session. 	<ul style="list-style-type: none"> We did not use because was too much time, participants lost the concentration and it is tired to them
	<ul style="list-style-type: none"> [17] DEVAN method to work directly with children with DS. 	<ul style="list-style-type: none"> We tried to apply but was too complicated and was focused only to childrens
	<ul style="list-style-type: none"> [19] Submission of evaluation criteria, Signature of anonymity and confidentiality 	<ul style="list-style-type: none"> We aplyed this suggestion because was required by the Tutors.
Usatestdown	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> Take the participants with the same level of disability. 	<ul style="list-style-type: none"> We selected participats taking on count only age but they had different skills levels and what might be simple for one may be more complex for others.
	<ul style="list-style-type: none"> The tasks should not have a high complex level. The difficult scale should be of 1 to 3, where 1 is esaier , the thask should take 1. 	<ul style="list-style-type: none"> The participants showed a problem to reach the normal objectives if the task is too complex, it means the level must to start on 1 and if it is necessary increase the task difficut level but the higerst level is not recommended
	<ul style="list-style-type: none"> The session should be done in 10 minute sessions for each person, 	<ul style="list-style-type: none"> The participantes had a problem to keep the attention for more than 10 minutes. because they will get tired easlily. This point will allow the individual to evaluate the application with curiosity without getting overwhelmed or bored.
	<ul style="list-style-type: none"> Do not limit to the participants the time. 	The participants will stress and become confused if they have a time limit for the task. We could see on the sessions that the participants were getting afraid with the topic task limited with time. They could feel nervous.

5.2 Write the instructions

Table 12: Usatestdown: Write the instructions

USATESTDOWN		
Step :	Write the instructions	
Definition	Specify the instructions given to the users (oral, written, or both) to perform each task	
Bibliography Research	NO CONTRIBUTIONS	
Usatestdown	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> Use only the oral way to explain the task to the participants 	<ul style="list-style-type: none"> Before completing the test, it is necessary to give a presentation to the persons who will participate with the evaluation. They should interact with the aplication on a training way. The Facilitator should explain the project objectives on a slowly way
	<ul style="list-style-type: none"> Ask the participant if he/she is a willing partipant. 	<ul style="list-style-type: none"> Evaluator must to ask if the participant would like to participate even if the Tutor recomeded this participant. The willingness of the participants to participate is very important to have real dates. They could feel pushed for the Tutors and the tutor could not realize about it.
	<ul style="list-style-type: none"> Speak slowly and repeat the task 	<ul style="list-style-type: none"> Speak slowly and repeat if is necessary, participants use to lost the attention easily
	<ul style="list-style-type: none"> Perform the sesion with the participant individualy. 	<ul style="list-style-type: none"> Parrticipants need to have extra attention and the evaluator should put all his attention on the participant to be sure he understood the task.

5.3 Define the test plan

Table 13: Usatestdown: Define the test plan

USATESTDOWN		
<i>Step :</i>	<i>Define the test plan</i>	
<i>Definition</i>	It is necessary to specify the protocol with alternative activities, such as, welcome, interview preview, completing the tasks by observing the user, satisfaction questionnaire, personal interview to collect qualitative information, etc. It is recommended to write an introductory commentary to express a welcome to the users. It is necessary, as part of these instructions, to collect the data needed by the users to complete the tasks.	
<i>Bibliography Research :</i>	BIBLIOGRAPHY SUGESTION	REAL EXPERIENCE
	[20] The evaluation was done on the users on a one-to-one basis.	It was necessary evaluate every participant on a individual way otherwise they lost the concentration. They need many attention and support.
<i>Usatestdown</i>	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> The demographic survey must have only general information not lasts names. 	<ul style="list-style-type: none"> It is necessary to complete a demographic survey including name, age, gender, experience with mobile devices. It must to have only general information, even it could use a fake name because family, tutors and participants are so reserved wit information that it would allow them to be identifier. Never take last names. Don't push the persons to answer if they don't want, it may generate a bad atmosphere to work. This is like the test that was used in the workshops and can be found in Annex 1. It should be completed by the evaluator.
	<ul style="list-style-type: none"> Applly the User satisfaction survey only with 3 answers . USATESRDOWN tets is recommended. 	<ul style="list-style-type: none"> It is important to prepare a user satisfaction survey with a scale of no greater than 3 categories and, if possible, with graphics of faces (happy, neutral, sad). We tried to work with 5 answers but it was confusing to the participants. Additionally we had a meeting with the students psicologist supervisors and they also recommended only 3 levels
	<ul style="list-style-type: none"> Don't use documents with text to the participants. If you need to give documents it is better with graphics or don't not use documents 	<ul style="list-style-type: none"> Avoid providing documents with long text to the persons who will participate in the test Generally, persons with DS have vision problems and it is taxing for them to read and speak. It is recommended that the instructions be given verbally and in a graphic form that is simple, allowing them to understand the information. In the first workshop session , we wrote test to explain to the participants the steps that they should follow but they had probles to read or pronounce and to understand. It si better to avoid
	<ul style="list-style-type: none"> Write short questions only with 3 answers 	<ul style="list-style-type: none"> Write short questions only with 3 answers (not, may be, yes), in this specific order because if they find fist the option YES they will not read the rest of te options. If you want to ask about quantity answer posibility is (few, many, too many). It means only 3 answers.
	<ul style="list-style-type: none"> Write questions focused on the applications and try to make understand the participants how important is to say the truth 	<ul style="list-style-type: none"> The participants are likely to provide friendly answers whether or not they like the application because they tend to be friendly persons. Persons with Down Syndrome are so friendly and they like to make friends, this is their normal behaviour, we had seeral cases with participants whoes answer the questionnaire saying the aplication was esy to understand, but when we analized the results with the log aplicaction the participants could not get success on the task or the success level was so low. Additionally when we asked Did you like the aplication ?' No body said NO, it showed us that it was not the rally truth. The wanted just to be friendly.
	<ul style="list-style-type: none"> Participation of the expert tutors is requiered 	<ul style="list-style-type: none"> The participation of the expert tutors is very important as their experience greatly contributed to the implementation of the test, following the guide. Additionally, it is critically important to include the expert tutors with the interaction with the participants'
	<ul style="list-style-type: none"> The sessions will be facilitated and observed by only one facilitator to one participant 	<ul style="list-style-type: none"> It is because on the workshops the participants were shy when they see many new persons in the room.

	<ul style="list-style-type: none"> Persons need to be encouraged to participate and facilitator should stress the value of the child's input and show appreciation and gratitude. 	<ul style="list-style-type: none"> This is a task that the facilitator should do. People with DS like to feel they are helping to another people. They will feel motivated.
	<ul style="list-style-type: none"> The educators of people with DS should be present at the moment of perform the session. 	<ul style="list-style-type: none"> This is very necessary on the session because the participants will feel sure and with a high confidence level.
	<ul style="list-style-type: none"> Give the participant the device for long time to execute the game (or task) task at home. Application should have a datalog to record. 	<ul style="list-style-type: none"> We suggest an evaluation stage where devices are given back to them to determine how much time is necessary for them to work independently from the tutors and then independent of the application and able to do the activity without the support of a tutor or the application.
	<ul style="list-style-type: none"> Apply at least 2 evaluation methods 	<ul style="list-style-type: none"> It means observational method to analyse the user behaviour and the tool should have a log or a way to evaluate if the task was completed successfully or not, additionally you should use a satisfaction questionnaire, in this case the proposal recommend the SUS Questionary adapted to persons with Down Syndrom.
	<ul style="list-style-type: none"> Establish objective metrics with a completion time for the task, error rate, etc. 	<ul style="list-style-type: none"> Performing the workshops we could notise the time is not a good parameter to take on count because when we said persons with down syndrome participants that they should do the task on a specific time, immediatly they were scared because they thought is an evaluation about how smart they are and it is obviously a big problem even to persons without down syndrome .
	<ul style="list-style-type: none"> The participants should have pre training about the application 	<ul style="list-style-type: none"> It mens tutor should teach to the participants how the application work and the should interact par minutes with the application before the real test.
	<ul style="list-style-type: none"> Establish subjective metrics such as sucess, frustration, satisfaction, etc. 	<ul style="list-style-type: none"> While the participants are using the application. Success: Defined as the completion of a task done correctly and without help. Satisfaction: When the user gets the correct result easily, when the user shows happiness when interacting with the application, etc. Frustration: When the user has problems answering a tutor's questions, when the user gives an incorrect answer, when the user doesn't understand the process, etc.
	<ul style="list-style-type: none"> Schedule a break halfway 	<ul style="list-style-type: none"> It means , through the test session and remind the participants that they can stop at anytime. Don't push participants
	<ul style="list-style-type: none"> Record videos about the participants sessions but without faces. 	<ul style="list-style-type: none"> To evaluate Success, frustration, satisfaction you could use the videos recorder during the workshop or session or at te same time the tutor could take on count measure of every tasks , how many times the participant showed this reaction.
	<ul style="list-style-type: none"> Hold a meeting prior to executing the test 	<ul style="list-style-type: none"> Because it is very important to break the ice with the participant so that they feel safe and trusting of the process at the workshop, This is a way to make friendly and relaxin the work enviroment
	<ul style="list-style-type: none"> Observation method needs facilitator to record all children action, behaviours and facial expressions while observing children playing the game. 	<ul style="list-style-type: none"> In order to guide facilitator, an observation checklist is needed to analyse the participants behavoiur.
	<ul style="list-style-type: none"> The evaluator should evaluate every participant on a individual way . 	<ul style="list-style-type: none"> The evaluation needs to catch the participat attention, until to be sure the participant understood the task . Some times is necessary ask the participants to see the evaluator eyes otherwise they focus their attention on another thing they could see.
	<ul style="list-style-type: none"> Do not use the technique of "thinking out loud" 	<ul style="list-style-type: none"> Because the majority of the participants have difficulty expressing themselves
Docu ments	<p>The documents should be printed</p> <ul style="list-style-type: none"> -UsatestDown Demographic Questionary (Annexes) -UsatestDown SUS Questionary adapted to persons with Down Syndrome focused -UsatestDown SUS Questionary to Tutors focused in the whole process (Annexes) 	

5.4 Run the pilot test

Table 14: Usatestdown: Run the pilot test

USATESTDOWN		
<i>Step :</i>	<i>Run the pilot test</i>	
<i>Definition</i>	Execute the test protocol using the welcome, the written instructions, completing the observations, measuring times, completing the interviews, etc. in order to analyze if the proposed process functions as expected. In the case that it is not, it should be written as the protocol describe.	
<i>Bibliography Research :</i>	BIBLIOGRAPHY SUGESTION	REAL EXPERIENCE
	[19] Involve field evaluation with a larger and more diverse user group to assess the extent	This recommendation does not define the number . In our case we recommend at least 10 users.
	[6] Don't use passwords	We never used passwords
<i>Usatestdown</i>	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	• Record on video the interaction of the person with the mobile device	• It is recommended and very important in order to qualitatively evaluate their interaction with it when we review the videos. Be careful not to film faces and to obtain authorization in order to complete this point.
	• Ask for parents permission if it is necessary to film faces,	• The evaluator should be really polite because this is a sensitive topic. Some times is necessary the fathers autorization to record participants faces .
	• To the pilot test is necessary only one participant	• It is recommended that the pilot test is done through a small samples but in this case it means only one person because the evaluator will notice with the second participant the mistake will be the same. This will allow a definition of the first process errors without needing to involve all of the participants in the process, repeating the same error.
	• The tutors or professors should be present on the evaluation session.	• During the entire evaluation process or user participation, it is necessary that the tutors or professors with whom they are interacting on a daily basis are present and provide a sense of support as we could see in the sessions we performed
	• The facilitator will sit next to the child during the session to take note and solve participant questions	• The facilitator will sit next to the child during the session and his/her role is to fill in the observation form while interacting with the child to make them feel at ease. We rocomend the participant doesnot write himself the questionnaire. They use to have problems to write or read.
	• Two questionnaires need to be assessed, they are proposed for USATESTDOWN	• There are two questionnaires that need to be assessed which are demographic questionnaire and post task questionnaire but we propose specifics questionaies adapted to participants with Down Syndrome. We designed this questionaries with Special Psicologist whoes work every day with those participants , taking as base the SUS questionarie.
	• Post task questionnaire will be conducted right after each test session with the help from the Tutor or Parent	• Post task questionnaire will be conducted right after each test session with the help from the Tutor or Parent. Is is neceary make the testas soon as posible , not the next day because participants forget what they did. Facilitador should be close just in case the participant have a dubt.
	• Take note of the times when the participant asks for help.	• The evaluator mus to take note of the times when the participant asks for help.
	• Take a pause	• Schedule a break halfway through the test session and remind the participants that they can stop at anytime. Don't push to the students to do everything at the same time.
	• Use simple words when directing the participant	• When you explain to the participants the tasks, process, objectives etc, you should use a easy vocabulary, people with down syndrome showed they don't understand difficult words and they lost attention
	• Speak slowly slow and some times is necessary repeat the same idea	• Because the participats have a problem to concentrate their attention.
<i>Documents</i>	The documents should be applied -UsatestDown Demographic Questionary (Annexes) -UsatestDown SUS Questionary adapted to persons with Down Syndrome focused in the whole process -UsatestDown SUS Questionary to Tutors focused in the whole process (Annexes)	

5.5 Refine the test plan after analysing the results of the pilot tests.

Table 15: Usatestdown: Refine the test plan after analyzing the results of the pilot tests.

USATESTDOWN		
<i>Step :</i>	<i>Refine the test plan after analysing the results of the pilot tests.</i>	
<i>Definition</i>	Once analyzing the results of the pilot test, modifications may be made to the protocol, instructions, task data, task sequencing, interview questions, etc., if necessary.	
<i>Bibliography Research:</i>	NO CONTRIBUTIONS	
<i>Usatestdown</i>	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> • Validate the tasks what the evaluator would you like to change with the Tutors 	<ul style="list-style-type: none"> • If an error is encountered in the test pilot, it is necessary to make an immediate change to the plan and execute a second session but it should be approved by the DS persons Tutors because evaluator could consider additional actions as normal but Tutors could have another point of view
	<ul style="list-style-type: none"> • Write the changes will be done and explain the reason why it was made. 	<ul style="list-style-type: none"> • It could help to understand the DS persons behaviour with the application. It should take on count to make inform with the final analysis
	It is required a formal document signed by the participants tutor to make changes.	<ul style="list-style-type: none"> • If the initial Test Plan has to be changed, the evaluator should have a document signed by the Tutor with the approbation. With persons with Down syndrome is very important follow this rules.

5.6 Recruit participants

Table 16: Usatestdown: Recruit participants

USATESTDOWN		
<i>Step :</i>	<i>Recruit participants</i>	
<i>Definition</i>	Process to determine the type and number of participants needed for the usability tests.	
<i>Bibliography Research:</i>	BIBLIOGRAPHY SUGGESTION	REAL EXPERIENCE
	<ul style="list-style-type: none"> • [17] Recommended 4 participants ,[19] 6 participants [18] 3 or 5 to five participants • [24] work with two paediatricians, primary school teachers and 11 children with DS. • [6] Nielsen's study showed that a group of five users with different background, mixed gender and aged five to six years old , they were able to find about 80% of the findings in a system. 	<p>About number of participants We could see with smaller groups as 2 or 3 the difference between them is big because every student have special skills and the results did not define a clear behavior pattern</p> <p>In this specific case 6 persons was not enough.</p>
<i>Usatestdown</i>	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> • It is necessary that evaluators determine the cognitive disabilities level of the participants that they need , only the age is not a parameter to be on count. 	<ul style="list-style-type: none"> • To establish the tasks, it should be considered that the participants have different levels of abilities and what might be simple for one may be more complex for others.
	<ul style="list-style-type: none"> • There should be at least 10 participants to the evaluation 	<ul style="list-style-type: none"> • We could see with smaller groups as 2 or 3 the difference between them is big because every student have special skills and the results did not define a clear behavior pattern.
	<ul style="list-style-type: none"> • Send clear information about the session objective to the Technology Department, if they are interested on the research and they could find a transportation to the centre they would contact with the Directives explaining the advantages they could have. 	<ul style="list-style-type: none"> • The first step in recruiting young participants is sending information about the study to the places who are working with the participants profile that we want to work, it means, there are different kinds of Centres working with down Syndrome depending of the students levels skills.
	<ul style="list-style-type: none"> • Evaluator and Tutor should have a meeting to define the participant profile and tutor should recommend the participants profile to make the task that the evaluator proposed. 	<ul style="list-style-type: none"> • We should not only take on count the participants age, we should analyze what are the special skills that every participant have. In our case at the beginning we found participants with low mental disabilities

		and another so extrem. It makes not homogeneous group because we were not evaluating really the application, we were evaluating just the cognitive disabilities.
	<ul style="list-style-type: none"> • The facilitator should ask the participant if he/she wants to collaborate 	<ul style="list-style-type: none"> • Because the participant should be voluntier. We had a case with a participant who behaved on a rude way . We asked him if he want to participate and he did not want. It is the best way to evaluate because some times they feel pushed to contribute. The not willing participant could change the evaluation results
	<ul style="list-style-type: none"> • Don't push them to finish fast. 	<ul style="list-style-type: none"> •Participants neet to take their time, if they take more more time than the programated time just tell him , we will change to another activity on a game way , never mention time requiered to finish the task.
	<ul style="list-style-type: none"> •Place could be the same where the participants work 	<ul style="list-style-type: none"> •The place to evaluate the participants is the same place where they work or study , it means don't leave the participants to an enviroment they don't know, it make the participants feel unsafe and shy. Try to make the test sesion on a room inside the institution.

5.7 Run the test session

Table 17: Usatestdown: Run the test session

USATESTDOWN		
Step :	Run the test session	
Definition	This is the essential part of the evaluation because it is here that the usability evaluation is completed. (1) Welcome; (2) Ask the participants to carry out the tasks; (3a) If performance is measured, measure the times, (3b) If performance is not measured, interrupt the user to clarify their decisions; (4) Note the number of errors and other objective data; (5) Distribute a satisfaction questionnaire and complete a personal interview.	
Bibliography Research:	BIBLIOGRAPHY SUGESTION	REAL EXPERIENCE
	[20] The evaluation could be made on the users on a one-to-one basis.	We noticed this making the test sessions
	[19] Support to the participants	The suport to the participants was really important because when the participants can not perform something they have te dendance to be passive.
Usatestdown	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	•Do not complete the final test on the same day as the pilot testing	•Because the users will be tired and confused if the first pilot process failed.
	•Record everithing the participant make and all his body behavior	•This information is very important to make the cualitative analysis
	•Make a pause during the test	•Schedule a break halfway through the test session and remind the participants that they can stop at anytime. Take note of the times when the participant asks for help.
	•Use simple language to speak with the participants	•Use simple words when directing the participants. When you explain to the participants the tasks, process, objectives etc, you should use a esy vocabulary and you must to speak
	•Take note of the times when the participant asks for help.	•The evaluator mus to take note of the times when the participant asks for help and it will be showed on the final inform like a parameter to take on count on the evaluation
	•Speak slowly and repeat some times	•Speak slowly and some times is necessary repeat the same idea to the same participant and trying to make he or she pay attention , the participants had a problem to concentrate.
	•The participant who participated on the pilot test can participate on the real test	•It is recommended to execute the complete test from the beginning, including the first participant who collaborate on the pilot test because the participants do not memorize easily. It should also include the changes that were made to the test plan after the pilot.
	•Take nnot of the all participant behaviour	•Consider the reactions of the persons being evaluated for each of the tasks that they complete. It is very important to determine their satisfaction level and the improvements

		that could be made in the next version. We had different participants' reactions such as: smile, sad, angry, worried, thinking long time, complains, asks, etc. Every one means important information at the moment to evaluate.
	• Apply USATESTDOWN questionnaires proposed	• In general, we can see that it was necessary to adapt the SUS questionnaire for the persons with Down syndrome because it is a complex survey for participants. SUS was modified to evaluate the USATESTDOWN process, the guide, which was designed with the expert tutors who work with the participants daily.
	• Solve all of the questions that the user has during the process,	• It is important because otherwise the participants could not follow the next step, the answer could be on a discrete way to support the participant but it is necessary because they have limitations to memorize or understand easily. Don't let the participant alone.
	• Apply the questionnaires USATESTDOWN propose as soon as possible.	• After completing the usability test session with a down syndrome participant, facilitator needs to ask the participant to answer post task questionnaire that USATESTDOWN propose as soon as possible because the participants use to forget easily.
	• Ask the participant if he/she wants to collaborate	• Don't push the persons to participate if they don't want, it may generate a bad atmosphere to work.
	• Don't push the participant to answer a question	• Don't push the persons to answer if they don't want, it may generate a bad atmosphere to work.
Documents	-UsatestDown Demographic Questionary (Annexes) -UsatestDown SUS Questionary adapted to persons with Down Syndrome focused in the whole process -UsatestDown SUS Questionary to Tutors focused in the whole process (Annexes)	

5.8 Analyse the collected information

Table 18: Usatestdown: Analyse the collected information

USATESTDOWN		
Step :	<i>Analyse the collected information</i>	
Definition	Analyze the objective data (times, errors, etc.), the more subjective data (satisfaction questionnaire and interviews), and all of the data that contributes to understanding the behavior of the evaluated persons from the usability test. The objective is to identify usability problems and propose improvements.	
Bibliography Research:	BIBLIOGRAPHY SUGGESTION	REAL EXPERIENCE
	The authors J. G. Wishart proposed the DEVAN method[41].	We tried to apply but was too complicated and was focused only to childrens
Usatestdown	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	• Compare the qualitative with qualitative dates.	• It is important to analyze the data with all of the parameters collected from the persons who participated in the usability test. This involves qualitative content (logs) as well as quantitative (user reactions). It means the results could be different because persons with down syndrome have the tendency to be friendly and they will not answer the questions on an objective way, they would say the answer the evaluator would like to hear. For this reason the evaluator should conduct an analysis of the part that appeared qualitatively in the evaluation and the quantitative data results
	• Take specific attention on applications data logs to make Statistics, it will show the real result	• The tool should help the evaluation with logs which allow the evaluators analyze the participants' interactions with the mobile device for example, how many times the participant touch the screen, on which way, up, down, left, right etc. Finally we could analyze this information in a quantitative way.
	• The collected data by observing when the persons performing the task scenarios could be here	• The results should show the collected data during the test session, they are so important to analyze the qualitative part.

	<ul style="list-style-type: none"> •The index indicators should be clear explained 	<ul style="list-style-type: none"> •Index as : Success, Satisfaction and Frustration Rate per Task and Document should be analysed it shows the most information when the evaluator is working with people with down syndrome, they use to be so expressive with their face or body behave. This includes; the persons feeling, fun, ease of use and their satisfaction level towards the game.
--	---	--

5.9 Report results to the development team or management.

Table 19: Usatestdown: Report results to the development team or management.

USATESTDOWN		
Step:	<i>Report results to the development team or management.</i>	
Definition	Prepare a presentation or report to explain the usability problems that were encountered and how they can be improved.	
Bibliography Research:	NO CONTRIBUTIONS	
Usatestdown	USATESTDOWN GUIDELINE	GOT FROM EXPERIENCE
	<ul style="list-style-type: none"> • All people involve in the test sesion should be on the presentacionand give their own point of view and comments 	<ul style="list-style-type: none"> •The results presentation should be done to all of the members of the group, with a clear document
	<ul style="list-style-type: none"> •The document to send to the Design Develop department all the data logs information should be included. 	<ul style="list-style-type: none"> •The inform to the Desig an Develop department should include all the, and with the respective backups such us, aplicacion logs, recorded videos, questioanries, etc.
	<ul style="list-style-type: none"> •Make a short and easy to understand inform to the Directives people on the centre, they use to use the information to improve the education way 	<ul style="list-style-type: none"> •We did a inform with the most important points found on the test sesion because the centres use this information to improve the way to teach, but the inform must not have participants personal information because it must be private
	<ul style="list-style-type: none"> •Save the participants information on a confidential way after the test sesion. 	<ul style="list-style-type: none"> •Apply the Ethical Issues in Recruiting Participants, it means follow the rules that each centre have to manage the participant information with ethical process. We should safe in a private and confidential way the collected information.
	<ul style="list-style-type: none"> •Include the old reports if it there would be from another interation and make a comparision. 	<ul style="list-style-type: none"> •In the case that the results were not satisfactory, improvements to the system should be made and it should be executed again, following the USA TESTDOWN guide.

We wrote a paper with this proposal and it was accepted in the ICAIT 2017 Conference[25]. It will be published on March 2017.

6 Evaluación

The USATESTDOWN guideline has been evaluated with a set of experiments involving persons with Down syndrome in usability testing of mobile applications. The results of the evaluation have been to improve the guideline. The proposal guide is iterative, we will apply the guide using 4 tools and three Down syndrome centre. See **Figure 15**.

Figure 15: Working Method Overview

- **Process:** We evaluated the guide with four Phases called for the tool name we used. It is showed on the figure 15 :
 - Phase 1: AsissTask
 - Phase 2: Money Handlig
 - Phase 3: PICAA 2
 - Phase 4: EUREKA

We performed four workshops to evaluate the USATESTDOWN proposal as we show on the **Figure 16**. We used four tools focused on people with Down syndrome. It means 176 participants. We contacted three Special Centre (Prodis [33], Maria Corredentora [42], Apadema [43]) on Madrid and they allowed us to work there.

Applying USATESTDOWN			
TOOL	PLACE	PARTICIPANTS	AGE
ASSISTASK (UAMadrid)	Prodis Centre Madrid - Spain	4 M 7 F	21 – 28 years old
Money-handling App Fraunhofer Germany	Ma. Corredentora Centre . Madrid - Spain	10 M 10 F	7 – 8 years old
		9 M 9 F	10 – 11 years old
	Apadema Centre Madrid - Spain	17F	26 – 56 years old
PICAA 2	Apadema Centre Madrid - Spain	17F	26 – 56 years old
	Ma. Corredentora Centre . Madrid - Spain	19 M 19 F	7 – 11 years old
EUREKA	Apadema Centre Madrid - Spain	17F	26 – 56 years old
	Ma. Corredentora Centre . Madrid - Spain	19 M 19 F	7 – 11 years old

Figure 16: Applying USATESTDOWN

6.1 PHASE 1: Evaluating USATESTDOWN with ASSISTASK

This document describes the application of the guide in a real case in the PRODIS Centre in Madrid Spain [44] with the tool ASSISTASK [45].

Applying the usability testing guide USATESTDOWN, the evaluators can easily administer the usability test with applications on mobile devices for persons with Down syndrome in the different workshops following the different steps that the guide proposes. We proposed this test with specific activities in order to evaluate the USATESTDOWN Guide.

Below is an explanation of what was done in each step of the Guide.

6.1.1 Establish the tasks to be used in the usability tests.

- **Work group:** The work group is composed of two special tutors that work daily with persons who have Down syndrome, one usability evaluation expert, and one expert in the ASISST -TASK application.
- **Tool:** The tool is called ASISST -TASK and is executed on a mobile device, in the case of this evaluation, on a touch screen telephone. The application was designed to help persons with Down syndrome or other cognitive disabilities perform different activities. The analysis process and policy separation that were sent to PRODIS are implemented in this application. This tool functions by scanning a QR code and then describes each activity the user can choose from. When the user completes a task they may press the next button in the application to continue on to the next step.
- **Test environment:** The workshop was held in the employment center. We attempted to maintain a familiar and normal environment as much as possible for the participants.
- **Equipment:** 2 mobile touch screen telephones with Android systems were used. These phones had the ASISSTASK application installed and ran on the Tree version of Android. The application

allowed the participants to carry out the policy classification process. The telephones also had free software installed to record the interaction between the user and the screen.

- Logistics: Two cameras were used to film the external interactions of the users and determine, based on their gestures and comments their frustration and/or satisfaction with the application and the process.
- Measures: We took different evaluative measurements, such as qualitative and quantitative.
 - Ratio of tasks completed correctly/incorrectly (VIDEOS)
 - Number clicks to complete the task (CLICK LOGS)
 - Number of times the user expresses frustration/satisfaction (VIDEOS)
 - Number of executed instructions (LOG)
- Documents required for the test: Automobile policy documents, Globalis, Home, General survey, SUS survey, SUS focused on persons with Down syndrome.

6.1.2 Write the instructions that participants will be given to perform the usability test

- This test will evaluate the basic flow of ASISST -TASK – a mobile technology system that was specifically designed to assist persons with cognitive disabilities in their workplace. In order to evaluate this, we employed mobile devices for two different purposes: (1) to show the interactive guides that were adapted for the user, the task and the user's context; and (2) to locate and provide directions in indoor environments [45]. The following is a description of the basic flow of the application's use in a real case at PRODIS.
- Each participant will have 3 documents to use with the application.
- These documents will be the same for each participant.
- We will keep track of the time.
- The user should complete the described activities in accordance with the document that they receive.

6.1.3 Define the test plan

The test plan is a protocol stating activities like the welcome, pre-test interview, observed task performance by user, etc. We can see in **Figure 17**.

- Welcome: We will say a few words about the process and then welcome the participants to the workshop.
- Demographic Test, done via demographic survey: The survey contains general questions regarding name, age, gender, and if the participant has a telephone with a touch screen. The information is not used to identify the participants; it is simply collected as general information.
- Training: The participants are trained with the documents. We explain to them the tasks and how the application works. We observe their task performance and note the times when the user begins to play with the application, if the user shows satisfaction (defined as when the user easily gets a correct result, if they are happy when interacting with the application, etc.), frustration (defined as if the user has problems answering questions from the tutor, if the answer is wrong, or if the user states that they do not understand the process).
- Basic application flow: The application was adjusted for the activities in a department of the Special Employment Center for persons with Down syndrome in the PRODIS Centre in Madrid. The Mutual department's goal was to classify the different types of insurance policies. These policies are sent daily and workers with Down syndrome should be classified in these policies according to different validation parameters. For this workshop we chose 3 document types: Document 1: Auto Policy

(take a photo, identify the policy type, review that the papers are signed (NO), is there a personal document? (NO), is anything handwritten? (NO), staple, locate “cars without incidence,” final screen); Document 2: Globalis policy (take a photo, identify the policy type, review that the papers are signed (NO), located the Globalis signature, final screen), Document 3: home policy (take a photo, identify the type of policy, review that the papers are signed, is there a personal document? is anything handwritten? (YES), staple all of the documents, put it in the basket “home without incidence,” final screen)

- Satisfaction questionnaire: The SUS questionnaire for persons with Down syndrome is given with the help of an evaluator. The evaluator reads each question slowly and waits for the user to answer. The System Usability Scale (SUS) is a simple, ten-item Likert attitude scale that provides a global view of the user’s subjective assessments of usability. The usability of a system, as defined by the ISO standard ISO 9241, Part 11, can be measured by solely taking into account the context of the system use. It was modified to the needs of persons with Down syndrome. During a meeting with the experts at PRODIS, we analyzed the questions in the SUS questionnaire. It was determined that the questions were too difficult to understand for these users so we selected only 5 questions and adapted the text to this specific group. The new scale only has three answers: YES, MAYBE, NO. All of the modifications made were recommended by Down syndrome experts.
- Personal interview with Experts: The interview was given to 2 expert tutors who work with persons with Down syndrome daily.
- (What do you think about the evaluation process? Do you think that the proposal guide is focused on the real needs of persons with Down syndrome? How can we improve the process? On a scale of 1 to 5, what is the qualification of this process?)
- SUS USATESTDOWN. SUS has proven to be a valuable evaluation tool, being both robust and reliable. It correlates well with other subjective measures of usability [6]. In the second part of the Satisfaction Test, the Down syndrome experts have to answer a SUS questionnaire modified to focus on the evaluation process.

Figure 17: The test plan process

6.1.4 Run the pilot test to analyze whether the process works to plan

- Initially, the test was only given to one participant, who was observed to be a little shy.
- The cameras were poorly placed because we were able to see part of the participant's face in the recording.
- The user seemed a little confused with the short explanation of the application.

6.1.5 Refine the test plan after analyzing the results of the pilot tests

- After the shyness of the first participant we decided to involve the tutors and have them work in parallel during the execution of the tasks.
- We modified the camera positions so that we were not filming the faces of the participants.
- We increased the time taken and detail given in the explanation given to each user.

6.1.6 Recruit participants

The guide proposes in this step to involve usability experts, down syndrome experts, down syndrome tutors to define the Down syndrome participant profile. In the Prodis case after determining the population group of interest and the required number of participants, participants were selected by the experts in the PRODIS Centre. Each tester ideally evaluates approximately 10 or more participants, if possible. The workshop was done with 7 women and 4 men between the ages of 21 and 28 years old. The participants did not have prior experience handling the insurance policies selection process.

6.1.7 Run the test session

With the test pilot experience, we applied the corrections to the test. After we did this, we continued performing the modified test with the rest of the participants.

6.1.8 Analyze the collected objective (times, number of errors, etc.) and subjective (satisfaction questionnaires) data

We revised the videos and annotations and created a document with the collected data. The Mutual department's goal was to classify the different types of insurance policies. These policies are sent daily and workers with Down syndrome should be classified in these policies according to different validation parameters. For this workshop we chose 3 document types with different flow options. With this process we have the following analysis .

6.1.8.1 Analysis of success, satisfaction and frustration

To analyze these characteristics, both videos from external cameras and recordings from the application that was installed on the mobile phones to record the user's interaction with the application were used.

Success: Defined as the completion of a task done correctly and without help.

Satisfaction: When the user gets the correct result easily, when the user shows happiness when interacting with the application, etc.

Frustration: When the user has problems answering a tutor's questions, when the user gives an incorrect answer, when the user doesn't understand the process, etc.

For document 1, 11 users participated and there were 8 classification tasks. For each task, the users were evaluated on success, satisfaction and frustration. The analysis was done by multiplying the number of users by the total number of tasks, which makes 88 tasks for document 1. The same process was done for document 2, which had 10 participants and 5 tasks for that document, for a total of 50 tasks for the document. Document 3 had 9 participants and 8 tasks, for a total of 72 tasks. Table 20.

Table 20: Analysis of success, satisfaction and frustration

C	Satisfaction	Frustration	Tasks* Document	Document Number	Users	Tasks *Users* Documents
85	103	198	8	1	11	88
48	47	131	5	2	10	50
66	77	153	8	3	9	72
199	227	482	21		30	210

Analyzing the information that was gathered shows that the rate of success of the users was 96%, which indicates to us that the majority of the tasks were done correctly by the users in Document 1. In Document 2, we also observed a success rate of 96%. Finally, in Document 3, there was a success rate of 94%. In general, the three documents had a 94% rate of success that shows that the participants had a total capacity to execute the application programs.

Satisfaction could not be measured as a percentage, so we took the total of the measures of satisfaction and divided them by the number of total tasks. Document 1 showed that there was a 1.2 satisfaction per task, which indicates that for each task that was completed correctly, the user expressed more than one indication of satisfaction. The frustration percentage for the same document was higher than that of satisfaction as, for each task, there were slightly more than 2 indications of frustration to perform the task until the users finally were able to complete it. For Document 2 there was almost an indication of satisfaction for each completed task. The percentage of frustration was the highest out of all of the three documents, with 2.6 indications of frustration for each completed task. For Document 3, there were 1.1 indications of satisfaction when performing a task and 2.1 times that the users indicated some form of frustration. In general, there was a success rate of 94.7%, which is incredibly high for the participants that were performing the tasks. Total user satisfaction upon completing a task was 1.08, which indicates that, in general, the users were satisfied with the tasks they had completed. The rate of frustration was high at 2.2 times that the participants indicated some form of frustration for each task. Given that it was the first time that the users had interacted with the application and did not have prior experience with the process of organizing policies, it is logical to think that the rate of complexity or frustration would be high. What is important is that the tasks were correctly completed, though it is important to remember that it took them longer to complete them. Table 21 and Fig 18.

Table 21: Success Rate per Document

DOCUMENT NUMBER	SUCCESS RATE	TASK SATISFACTION	FRUSTRATION PER TASK
1	96.5909091	1.17045455	2.25
2	96	0.94	2.62
3	91.6666667	1.06944444	2.125
TOTAL	94.7619048	1.08095238	2.2952381

Figure 18: Success Rate per Document

6.1.8.2 Time analysis

Thanks to the registries recorded by AssisT-Task, we were able to analyse users' performance by means of the time needed for each document. In **¡Error! No se encuentra el origen de la referencia.** we plotted the time needed by each user to process each document. As can be noticed, document 2 took less time than the others for all users. This is motivated by the fact that users had to do less steps to process it. However, the most important conclusion that the plot reflects is the common difference between documents 1 and 3. As can be seen, most users required less time to process document 3. This difference is motivated by the knowledge acquisition, both in doing the task and working with the application, since both tasks consisted on the same number of steps and there were no other fundamental differences in the execution. Fig 19.

Figure 19: Total time needed to process each document, expressed in seconds

6.1.8.3 SUS Usability Satisfaction Test Evaluation

The SUS Usability Satisfaction Test scale was modified based on the recommendations made by Down syndrome specialists. The test was modified to have 5 questions rewritten in a simpler, clear way. The scale used to analyze the data is based on 3 levels (Yes = 2, Maybe = 1, No = 0). The evaluation process of the

test is the same as that of the standard SUS test. The analysis process of the answers begins by adding all of the answers for the same questions, with a total of 22 questions. This is considered to be 100% and used to determine percentages. Table 22.

Table 22: SUS Questionnaire for persons with Down syndrome

SUS QUESTIONNAIRE	YES	MAYBE	NO
1. Would you use this application to work?			
2. Do you think the application is difficult to use?			
3. Do you think the application is easy to use?			
4. Do you think help is necessary in order to use the application?			
5. Do you think the application would be useful for your work colleagues?			

The modified SUS test was given to the 11 workshop participants and the results for questions 1, 2, 3, 4 and 5 are acceptable according to the SUS scale but question 4 had a low percentage given that the participants needed help at the beginning to work with the application. This is to be expected since none of them had previous experience working with the tool or with classifying policies. Table 23 and Fig 20.

Table 23: Modified SUS Analysis

Question 1	Question 2	Question 3	Question 4	Question 5
100.0	72.7	68.2	59.1	100.0

Figure 20: Questions and acceptance percentages of SUS participants

6.1.8.4 Questionnaire for Down syndrome experts

Interviews were done with 2 Down syndrome experts who work daily at the PRODIS Centre. Two questionnaires were used, the first with general questions and the second was a SUS satisfaction test that was modified to specifically evaluate their satisfaction with the guide.

6.1.8.4.1 General Questionnaire

The general questionnaire was given to the expert tutors, using open questions in accordance with the evaluation process (i.e., it was focused on USATESTDOWN in order to determine if the proposed guide is useful and applicable in our case study). See Table 24.

Table 24: General Questionnaire

Question	Expert 1	Expert 2
What do you think about the evaluation process?	-The prior test interview with the usability and tutor experts at PRODIS was very productive. -The proposal was corrected by adjusting it to the participants' reality. -Well prepared.	-Good focus and flexibility. -In this area, the common room. -Well equipped. -Well-prepared workshop. -Adequate coordination given the means and resources available.
Do you think that the proposal guide is focused on the real needs of persons with Down syndrome?	Yes, since the process is done in an individualized manner and with the support of expert mediators, the participants don't react in the same way as they would with strangers.	Yes, it can be clearly seen that the participants are motivated, autonomous, and comfortable. At the end, they expressed that they enjoyed the process.
How can we improve the process?	-Less evaluation subjects or -Increase the user's interaction time -Increase the number of usability experts so the evaluation can be conducted in parallel. -The interviews with the participants should be done immediately after because they have short-term memory.	-Increase the time for each participant. -Two participants in the same room is not optimal. -The third policy was done the best, indicating that they are more agile with the application after some time to experiment with it. -Avoid having the participants wait for the mediator or tutor to say something before they begin an action.
On a scale of 1 to 5, how do you rate this process and why?	4 because the evaluation time can be improved. -In general, the evaluation process is very useful and can adapt to the needs and limitations of the participants.	4.5 because the initial coordination to modify the methodology and tool to this field of study required some time to complete, with previous attempts, reviewing the documents, reviewing the process, etc.

Table 5 shows the different comments from the expert tutors at the PRODIS Centre, of which, the majority are positive. While they recommended improving the time, these suggestions will be taken into account in our report and used to improve the guide.

6.1.8.4.2 SUS Applied to USATESTDOWN Tutors Experts.

In the second part of the Satisfaction test, the Down syndrome experts answered a standard SUS questionnaire, focusing on the evaluation process. The test can be found in APPENDIX 1. The answers to the 10 questions from the SUS, which was adapted for the USATESTDOWN guide can be found in Table 6. The tutors that worked on the entire process were 2. The analysis shows that, following the SUS evaluation scale, all of the questions were accepted at a rate of 68% or more. The questions with the lowest score (75) were question 3 (I thought the guide was easy to use) and question 4 (I think that I would need the support of an expert to be able to use this system). However, these two questions scored above the acceptable limit of the scale and it can be concluded that the guide meets the needs of persons with Down syndrome. Fig.25.

Table 25: SUS, modified for USATESTDOWN

QUESTION	EXPERT 1	EXPERT 2	TOTAL	RESULT
1	4	4	8	100
2	4	3	7	87.5
3	3	3	6	75
4	3	3	6	75
5	4	3	7	87.5
6	4	4	8	100
7	4	4	8	100
8	4	3	7	87.5
9	4	3	7	87.5
10	4	4	8	100

Figure 21: Acceptance rate percentages of the questions for the tutors

6.1.9 Report results to the development team or management.

We will show the result report to the development group team and the Direction Department.

6.2 Evaluating USATESTDOWN through Money Handling tool

We apply the same process that on the 6.1 with special changes focused on the new application functions . We describe the process to evaluate the usability of Money Handling Tool App with the USATESTDOWN guide inn the Figure 22.

Figure 22: Workshop process for testing MHT App with USATESTDOWN

At the beginning, we individually explained to the participants the purpose and process and requested their consent on testing the game. They agreed with the option to play with a mobile device. Additionally, we asked if they agree that their hands are recorded, to which all participants agreed. We asked the participants to fill in a general questionnaire about age, name, and previous experience with touch screen devices. Then, we worked with the participants individually and explained how the application works in an easy way. This helps the participants to understand a real example with a shop that the students know.

For every device we conducted three tries. This means each participant had to buy three groceries using the MHT App. There was no specific time limit to finish the task, because they may feel uncomfortable or forced to end the task, which could change the evaluations results. After the participants did the six tries to buy (three for every device), we posed a satisfaction questionnaire, asking which device they prefer, to play with the application.

During the personal interview we asked the participants if they think the application is useful for everyone. Another question was if they think it is easy to use or not.

Maria Corredentora Centre is a special education centre in Madrid for children with DS [1]. The centre has four education levels: Early Childhood education, Basic stage education, DVA stage education, and professional programs. There are children with DS, between 5 and 16 years old. Each group has finished general skill development. We selected our evaluation group taking into account their age as our main parameter. From the remaining group we randomly chose the participants.

This section describes the execution of the previously described USATESTDOWN method applied on the MHT App.

We divided the evaluation process in two phases according to our two evaluation sessions. The goal was to evaluate the usability of the MHT App version 2.

As test systems we used two mobile devices - a mobile phone with 4.7" screen and a tablet 10.1" screen, with Android 6, as shown in Figure 22.

Figure 23: Mobile devices used for testing usability depending on the device size; Tablet with 22.2" and mobile phone with 4.7" running Android 6.

In the first phase, we used both the mobile phone and the tablet. The participants were 20 children between 7 and 8 years old. In this case it was not necessary to ask about touch screen experience, as the Maria Corredentora Centre has an official tool called PICCA 2 which works on a tablet. Using this App the centre teaches the children at least twice per week. We followed the process of USATESTDOWN, as described before.

6.2.1 Evaluation of the Money-handling Training App – comparison of different age groups

We wanted to know if the difficulty of using the MHT App depends on the user's age. We followed the same USATESTDOWN process as described before. For this we used two user groups of different age spans. The first user group was at the Maria Corredentora Center, where we worked with 18 children between 10 and 11 years. The second user group was located in the Apadema Centre, which is an Occupational Center that hosts 44 women [2]. The participants were 17 women with DS, between 26 and 56 years old.

The satisfaction questionnaire result from the evaluation comparing the device size indicated that the children prefer to work with the tablet and not with the cell phone. The main reason was the MHT App interface, as the buttons and icons are too small, to be properly displayed on a mobile phone screen. It was a major concern for DS users, as many of them have vision problems. For this reason we used a 10.1" tablet as testing device, see Figure 13.

First findings show that there is a difference between children and adults people with DS, managing the MHT App. Children needed more support on mathematic logic. More explanation of the basic operations was required for the children, compared to the adults. However, this had only a minor impact on the success rate.

Figure 24: Money Handling with Adults with Down Syndrome

• Conclusions

At both centres, 82% of the participants answered that the application is useful in the real word and they would like to play with it. Additionally, 91% of the participants answered that the application is complicated to manage. The application therefore is in principle suitable to teach people with DS, and should be extended.

6.3 Evaluating USATESTDOWN through PICCA2

We applied the same process described on the point 6.1. with different tasks

PICAA 2 is a platform to design educational activities for users with education special needs. Its aim is helping in the design of learning activities, which can be personalized by teachers at content and user interface levels. In this section we will analyze the learning process to determine the characteristics of users and activities. Then we will expose the main features of the platform and finally how to treat personalization and adaptation [46]. See Figure 25.

We applied this workshop in two special Education Centre to people with Down Syndrome

- Apadema Centre with 17 woman participants between 26 and 56 years old.
- Maria Corredentora Centre with 19 children and 19 girls between 7 and 11 years old.

It means 55 participants who collaborated to evaluate this application with USATESTDOWN.

Figure 25: PICAA 2 with people with Down Syndrome

6.4 Evaluating USATESTDOWN through EURECA

Eureca is an application focused on people with down syndrome to help them to play or increase their skills to manage a mobile device. The application have games, puzzles, etc. The teacher, tutor or parents can design activities to play. See Figure 26

We applied this workshop in two special education Centre to people with Down syndrome

- Apadema Centre with 17 woman participants between 26 and 56 years old.
- Maria Corredentora Centre with 19 children and 19 girls between 7 and 11 years old.

It means 55 participants who collaborated to evaluate this application with USATESTDOWN.

Figure 26: EUREKA with people with Down Syndrome

6.5 Conclusions from the evaluation of USATESTDOWN

In order to the main conclusions from the evaluation of USATESTDOWN we have that in the Phase 1 we found the most important findings it means 70 % of the changes added . The 20 % was in the phase 2 ant with the evaluation on phase 3 and 4 we did not find big additional contributions. The process was performed without bugs. It means the Guide is acceptable.

6.5.1 Conclusions Phase 1

In general, we can see that it was necessary to adapt the SUS questionnaire for the persons with Down syndrome because it is a complex survey for participants. The test results were overwhelmingly positive and

participants commented that they had enjoyed both the application and the process, which can also be confirmed in the recorded videos.

SUS was modified not to the application or evaluation system, but to the guide, which was successfully applied to the expert tutors who work with the participants daily. Evaluating the two expert tutors, the results were completely passed in accordance with the SUS evaluation scale, which has a minimum limit of 68, as all of our questions scored above 70%.

In general, it is clear that the guide is viable and can be successfully used and modified to the needs of persons with Down syndrome, with this as an example of a real-world success. It was also evaluated by the expert tutors as part of this process, which was a great help and supported the adaptation of the guide.

The participation of the expert tutors was very important as their experience greatly contributed to the implementation of the test, following the guide. Additionally, it is critically important to include the expert tutors with the interaction with the participants' interactions with the application to create a comfortable and familiar environment so the participants feel safe and trust the process as they are asked questions or doing a task.

We can say that a negative factor of this evaluation was the time it required to carry out the pilot plan with more than one participant, because using the time parameter set by the first participant to force the second participant to complete the task in the same amount of time will not always produce the same cognitive or memory coefficients. We recommend that times are not as strict and participants are able to work with as much flexibility as possible.

It should also be noted that the participants quickly forget the process. We suggest an evaluation stage where devices are given back to them to determine how much time is necessary for them to work independently from the tutors and then independent of the application and able to do the activity without the support of a tutor or the application.

We applied this finding on the USATESTDOWN guide.

6.5.2 CONCLUSIONS PHASE 2

After we applied the findings in evaluation Phase 1, we evaluated again the guide with another tool and another participants and centre to discard the participants learning process. We found the follow results.

We asked every participant about their device preference. 95 % of participants (20 children) answered that they prefer the tablet for interaction. This was also confirmed with the participant's behaviour when they took the mobile telephone close to their face, because it was difficult to see the different buttons and icons on the screen. Most persons with DS have vision problems, which is a real limitation for applications that have too many icons.

Furthermore, we analysed the videos looking for the success or failure rate, when buying the product. If the participant manages to buy at least two products, it was considered a successful try. At Maria Corredentora Centre we found only one student that succeeded. At Apadema Centre the analysis showed that one participant bought 2 products and another participant bought 3 products, leading to two successful executions of the task.

We evaluated the guide with Children's and with adults with Down syndrome. We found that the success factor was the same with the two groups, it means with this application the mathematic complexity level was the same to the both groups. We supposed in adults there would be a big difference. We should analysed the mental age not the year age.

We used two touch screen devices with the same application (Mobil cell phone ant tablet 11"), We found that the students prefer to use the tablet to work because they have vision problems. We concluded usability is proportional to the mobile device size.

First findings show that there is a difference between children and adults people with DS, managing the MHT App.

We added this findings to USATETDOWNN guide

6.5.3 Conclusions Phase 3

We found it is necessary to maket differents repots to every area who participate on the work shop. The report to the Special Centre should be written with no technical expressions. It must to be easy to understand. Don't write students names in the report.

To the Software Develop department the report should be objective and concrete saying if is it easy or not from the participants point of view.

6.5.4 Conclusions Phase 4

After the third test session USATESTDOW was complete but we did a fourth test session (Maria Corredetora, Apandema, EUREKA, 55 participants) and finally we analyzed the results we did not need fix anything more.

With this Evaluation Chapter we published one paper on the UCAMI 2016 Conference [39] . We wrote a JCR article in Journal of Ambient Intelligence and Smart Environments. Thematic Issue on Human-centred AmI: Cognitive Approaches, Reasoning and Learning. JAISE with collaboration of Fraunhofer Institute Germany . It was approved and it will be published on April 2017.

7 Conclusions and Results

7.1 Results

On the first phase we presented a review of research related to the usability testing of mobile applications including participants with Down syndrome. The purpose is to identify good usability testing practices and possible guidelines for this process when participants are people with this cognitive disability. These practices and guidelines should account for their specific impairments. We applied document analysis techniques to searches of scientific databases. The results were filtered considering how well they matched the research topic. We processed and reported the classified and summarized results. The main findings of this literature review is that mobile applications usability testing including people with Down syndrome is an issue that has not been comprehensively investigated. While there is some related research, this is incomplete, and there is no single proposal that takes on board all the issues that could be taken into account. Consequently, we propose to develop guidelines on the usability testing process involving participants with Down syndrome. We got a publication. "Usability Testing Process With People With Down Syndrome Interacting With Mobile Applications: A Literature Review"[14].

On the second Phase, at the beginning we were working with Cognitive disabilities (People with Down syndrome and Elderly people). For that reason we proposed a "Usability Evaluation Method for Mobile Applications for the Elderly : A Methodological Proposal" it was published by UCAMI 2014 [47] We performed an evaluation session test to 101 elderly people aged from 61 to 92 years to determine their ability to perform a number of gestures on multi- touch surfaces and the results were showed with the paper "Examining the Usability of Touch Screen Gestures for Elderly People" it was published by UCAMI 2016 [48]. After that we performed the evaluation test to 122 persons, they were 69 children and 53 adults with Down syndrome to determinate the skills, behave and how they interact with mobile devices. We got a Journal publication [27] and we wrote another paper called "Examining the Usability of Touch Screen Gestures for Adults with Down Syndrome" and we are waiting for the answer after 2 magazine corrections were made. With all this work we could compare and find the huge differences between people with Down syndrome and elderly people at the moment to perform applications with a touch screen mobile device. Then we decided separate this two population group and treat everyone like an individual group. After that we worked only with persons with Down syndrome because they showed better skills to perform mobile devices.

The third Phase merge the findings of the research bibliography and mainly the evaluation taking as base the formal Usability Testing Process as defined ISO on a USATESTDOWN guide . We wanted first to test the Guide before sent a paper because we needed to know we should make changes or not.

On the fourth phase or Evaluation, we did an exhaustive research, we applied this evaluation guide USATESTDOWN to 176 participants into 3 Centre on Madrid and with 4 tools. This was an iterative process. With the first test session (Prodis, ASSISTASK, 11 participants) we found the most part of mistakes. It was really productive to improve the USATESTDOWN proposal. With this work we got a paper called "Evaluation of a usability testing guide for mobile applications focused on people with Down

syndrome (USATESTDOWN) " it was published by UCAMI 2016 [39]. With the second test session (Maria Corredetora, Apandema, MoneyHandling, 55 participants) the evaluation to USATESTDOWN was almost complete, the most errors were fixed with the first evaluation. In the second test session the participant's age were different than in the first test session however USATESTDOWN worked as we expected. We wrote the paper "Assistive Apps for Activities of Daily Living supporting Persons with Down's Syndrome" to the Journal of Ambient Intelligence and Smart Environments like a Special Special Issue in cooperation with the Fraunhofer Institute Germany who developed the application Money Handling [49].We are waiting the Journal answer.

We found another small things to take on count to the next to test sessions (Maria Corredetora, Apandema, PICCA, 55 participants). After the third test session USATESTDOWN was complete but we did a fourth test session (Maria Corredetora, Apandema, EUREKA, 55 participants) and finally we analyzed the results we did not need fix anything more. The Tutors, parents, students were satisfied with the process.

The teamwork formed by participants Tutors, participants Parents and technical specialist was successful because each one contributed on their own expert area making USATESTDOWN a complete and strong guide on all sides. We collected all point of view from each professional and mainly from the participants who were glad collaborating.

In general, it is clear that the guide is viable and can be successfully used and modified to the needs of persons with Down syndrome, with this as an example of a real-world success. It was also evaluated by the expert tutors as part of this process, which was a great help and supported the adaptation of the guide.

7.2 Conclusiones

As I showed with this research, people with Down syndrome have a lot of skills to manage mobile and touch screens devices. Children's and Adults with DS could manage very the applications with the devices tested. They had a lot of motivation to work when It is a big strength to be used to help this special people to have a better stile of life. For example there are applications to show the persons the route how to arrive at home or how wear the clot making match the colours. Applications who help to remind the medicine on time or just to call someone to have help if they are on problems. Also we can find tools to help the persons with their task in the real work place as we did on Prodis Centre with the application Assistask or with Money handling application in Maria Corredetora and Apandema. Centres in Spain.

Only made the applications is not enough because developers don't take on count the special skills that this special persons have, the intention is help the people and it is great. The question is. Are we really helping this persons with the application?. After this research we could see it was not real. This is the point where USATESTDOWN help. One application with a good usability level for this persons is really important because they are afraid to use complicated applications. The down syndrome people fight every day to feel part of the "Normal Word" . If they find the application difficult to them and to us is easy (because we developed) it make feel them sad and it create a resistance level to use the application because it would make

them no smart as you wait . For this reason is very important to have a high usability level adapted to them special skills.

For example with Money Handley application developed by Fraunhofer Institute. It is a application developed specially to people with down syndrome like a small wallet with money to simulate the daily life of the persons when they go to the shopping to buy groceries. Apparently the application work well to the developers but when we applied UsatestDown guide we could find aspects they never take on count. The general evaluation results was: Money handling application is too complicated to this special people even when the application was developed focusing only on persons with down Syndrome . Now the developers are working in a new application version version.

The contribution with this research is work together with this important projects who try to help people with down syndrome but adapting the tool to the real skills that those persons have . UsatestDown will help with the usability testing guidelines to adapt this tools to real users and make this applications really helpful to the participants breaking the resistance barrier to use the application.

8 Future Work

The next line work will be analyzed the guide with participants and tools at home, it means the application must to record with a data log or screen recording the participants interaction at home and when the student is alone for a period for example one month, after a pre training process. I could show how the participants react or how they think without pressure, support and without guide.

May be interesting evaluate USATESTDOWN with gps applications, there are applications to show the participants how they could go to home, school or work. We applied the guide only in the participant workplace to make feel them sure. But we could analyze the participant's behavior outside their normal environment. We could find new topics to recommend and complete the guide.

We can follow the study with elderly people and create a specific USATESTELDERLY guide focused on Elderly people. This new guide could help to Elderly people to increase the use of mobile devices to help themselves with Mobil applications to make easy their life or solve healthy necessity problems as support technologies

8.1 Publications

8.1.1 JCR

JOURNALS DORIS CALIZ			
TITLE	JOURNAL	IF	STATE
Examining the Usability of Touch Screen Gestures for Adults with Down Syndrome	Journal of Ambient Intelligence and Humanized Computing.	IF (2015) = 1.462	
Assistive Apps for Activities of Daily Living supporting Persons with Down's Syndrome (Fraunhofer Institute Germany)	Journal of Ambient Intelligence and Smart Environments	IF (2016) = 0.707	APROVED
Examining the Usability of Touch Screen Gestures for Children with Down Syndrome	Interacting with Computers	IF (2014) = 1.268	APROVED
Creating TUIs Using RFID Sensors—A Case Study Based on the Literacy Process of Children with Down Syndrome	Sensors Journal	IF (2014) = 2.245	PUBLISHED

8.1.2 Conferences

CONFERENCES DORIS CALIZ			
PAPER	CONFERENCE	STATE	DATE
Usability Testing Process With People With Down Syndrome Intreacting With Mobile Applications: A Literature Review	International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 3, June 2016	PUBLISHED	June 2016
Examining the Usability of Touch Screen Gestures for Elderly People	10th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2016, (IWAAL & AmIHEALTH included)	PUBLISHED	2-Nov-16
Evaluation of a usability testing guide for mobile applications focused on people with Down syndrome (USATESTDOWN)	10th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2016, (IWAAL & AmIHEALTH included)	PUBLISHED	2-Nov-16
Usability Evaluation Method for Mobile Applications for the Elderly: A Methodological Proposal	10th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2014, (IWAAL & AmIHEALTH included)	PUBLISHED	2-Nov-14

Automatic Tagging As A Support Strategy For Creating Knowledge Maps	Society and Information Technologies: ICSIT 2017	WILL PUBLISHER	Feb-17
“USATESTDOWN” a proposal of a Usability testing guide for mobile applications focused on persons with Down syndrome.	International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 3, June 2017	APROBED	FEBRARY 2017

9 Bibliografía

- [1] C. Aal, T. Help, D. Syndrome, and L. B. Lives, “AAL 4 DS (Can AAL Technology Help People with Down Syndrome to Live Better Lives ?).”
- [2] J. Feng, J. Lazar, and L. Kumin, “Computer Usage by Children with Down Syndrome : Challenges and Future Research,” *Computer (Long. Beach. Calif.)*, vol. 2, no. 3, p. 13, 2010.
- [3] D. Rømen and D. Svanæs, “Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users,” *Univers. Access Inf. Soc.*, vol. 11, no. 4, pp. 375–385, 2012.
- [4] J. Feng, J. Lazar, L. Kumin, and a Ozok, “Computer Usage by Young Individuals with Down Syndrome: An Exploratory Study,” *Proc. 10th Int. ACM SIGACCESS Conf. Comput. Access.*, pp. 35–42, 2008.
- [5] J. Feng, J. Lazar, and L. Kumin, “Computer Usage by Young Individuals with Down Syndrome : An Exploratory Study,” pp. 35–41, 2008.
- [6] L. Kumin and J. Lazar, “A Usability Evaluation of Workplace-Related Tasks on a Multi-Touch Tablet Computer by Adults with Down Syndrome,” *J. Usability ...*, vol. 7, no. 4, pp. 118–142, 2012.
- [7] M. BRUNI, “Fine motor skills in children with Down syndrome: A guide for parents and professionals,” *Woodbine House*, 1998.
- [8] J. Nielsen, “Nielsen Norman Group,” *Nielsen Norman Gr.*, p. Articles, 1995.
- [9] I. Standard, “DRAFT INTERNATIONAL STANDARD ISO / FDIS,” vol. 2009, 2010.
- [10] F. Adebessin and P. H. Gelderblom, “Technical Report Usability and Accessibility Evaluation of the Digital Doorway.”
- [11] D. Hutchison and J. C. Mitchell, *HCI and Usability for Education and Work*. Torremolinos-Málaga, Spain: Springer, 2011.
- [12] S. D. Barnum, Carol M., *Usability Testing and Research*. Needham Heights, MA, USA ©2001: Allyn & Bacon, Inc, 2001.
- [13] a Kirijian, M. Myers, and S. Charland, “Web fun central: online learning tools for individuals with Down syndrome,” in ... *Users. Chichester, UK: John Wiley & ...*, 2007, p. 40.
- [14] D. Cáliz, L. Martínez, X. Alamán, C. Terán, and R. Cáliz, ““ U Sability Testing in Mobile Applications Involving People With Down Syndrome : a Literature Review ,”” *ICAIT 2016 Conf.*, vol. 8, no. 3, 2016.
- [15] J. Lazar, L. Kumin, and J. H. Feng, “Understanding the Computer Skills of Adult Expert Users with Down Syndrome : An Exploratory Study,” pp. 51–58, 2011.
- [16] J. W. Barbosa Chacón, J. C. Barbosa Herrera, and M. Rodrigue Villabona, “Revisión y análisis documental para estado del arte: Una propuesta metodológica desde el contexto de la sistematización de experiencias educativas,” *Investig. Bibl.*, vol. 27, no. 61, pp. 83–105, 2013.
- [17] I. Macedo and D. G. Trevisan, “A Method to Evaluate Disabled User Interaction : A Case Study with Down Syndrome Children,” *Univers. Access Human-Computer Interact. Des.*

Methods, Tools, Interact. Tech. eInclusion, pp. 50–58, 2013.

- [18] R. Ramli and H. B. Zaman, “Designing usability evaluation methodology framework of Augmented Reality basic reading courseware (AR BACA SindD) for Down Syndrome learner,” *Proc. 2011 Int. Conf. Electr. Eng. Informatics, ICEEI 2011*, no. July, 2011.
- [19] F. Adebessin, P. Kotzé, and H. Gelderblom, “The complementary role of two evaluation methods in the usability and accessibility evaluation of a non-standard system,” *Proc. 2010 Annu. Res. Conf. South African Inst. Comput. Sci. Inf. Technol. - SAICSIT '10*, pp. 1–11, 2010.
- [20] R. L. Yussof and H. Badioze Zaman, “Usability evaluation of multimedia courseware (MEL-SindD),” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5857 LNCS, pp. 337–343, 2009.
- [21] S. K. D. Mello, S. D. Craig, B. Gholson, S. Franklin, R. Picard, and A. C. Graesser, “Integrating Affect Sensors in an Intelligent Tutoring System,” *Affect. Interact. Comput. Affect. Loop Work. 2005 Int. Conf. Intell. User Interfaces*, pp. 7–13, 2005.
- [22] F. Alcantud, J. Coret, E. Jiménez, S. Márquez, F. Moreno, and J. Pérez, “Usability remote evaluation: METBA system,” *2012 15th Int. Conf. Interact. Collab. Learn. ICL 2012*, 2012.
- [23] P. T. Jaeger, “Assessing Section 508 compliance on federal e-government Web sites: A multi-method, user-centered evaluation of accessibility for persons with disabilities,” *Gov. Inf. Q.*, vol. 23, no. 2, pp. 169–190, 2006.
- [24] R. L. Yussof and T. N. S. T. Paris, “Reading Activities Using the Scaffolding in MEL-SindD for Down Syndrome Children,” *Procedia - Soc. Behav. Sci.*, vol. 35, no. December 2011, pp. 121–128, 2012.
- [25] D. Cáliz, L. Martínez, X. Alamán, C. Terán, and R. Cáliz, ““ U Sability Testing in Mobile Applications Involving People With Down Syndrome : a Literature Review ,”” *ICAIT 2016 Conf.*, 2016.
- [26] J. Jadan-guerrero, J. Jaen, D. Valencia, C. De Vera, U. D. C. Rica, and U. D. C. Rica, “Kiteracy : A Kit of Tangible Objects to Strengthen Literacy Skills in Children with Down Syndrome,” pp. 15–18, 2015.
- [27] J. Jadán-Guerrero, L. Guerrero, G. López, D. Cáliz, and J. Bravo, “Creating TUIs Using RFID Sensors—A Case Study Based on the Literacy Process of Children with Down Syndrome,” *Sensors*, vol. 15, no. 7, pp. 14845–14863, 2015.
- [28] J. Jadán-guerrero, G. López, and L. A. Guerrero, “Use of Tangible Interfaces to Support a Literacy System in Children with Intellectual Disabilities.”
- [29] V. Nacher, J. Jaen, E. Navarro, A. Catala, and P. González, “Multi-touch gestures for pre-kindergarten children,” *Int. J. Hum. Comput. Stud.*, vol. 73, pp. 37–51, 2015.
- [30] Asindown ORG, “Asindown Valencia,” 2017. [Online]. Available: <http://asindown.online/>. [Accessed: 10-Jan-2017].
- [31] “Maria Corredentora Centre,” 2016. [Online]. Available: <http://mariacorredentora.org/wpmcorg/>. [Accessed: 11-Nov-2016].
- [32] Apadema, “Apadema,” 2016. [Online]. Available: <http://www.apadema.es/>. [Accessed: 11-

Nov-2016].

- [33] Comunidad de Madrid, “Fundacion Prodis.” [Online]. Available: <http://www.fundacionprodis.org/>.
- [34] M. Schrepp, “GOMS analysis as a tool to investigate the usability of web units for disabled users,” *Univers. Access Inf. Soc.*, vol. 9, no. 1, pp. 77–86, 2010.
- [35] V. R. Basili, G. Caldiera, and H. D. Rombach, “The Goal Question Metric Approach,” in *Encyclopedia of Software Engineering*, vol. 2, Wiley, 1994, pp. 528–532.
- [36] R. Malak, M. Kotwicka, A. Krawczyk-wasielewska, and E. Mojs, “Motor skills , cognitive development and balance functions of children with Down syndrome,” vol. 20, no. 4, pp. 803–806, 2013.
- [37] J. G. Wishart, “The development of learning difficulties in children with Down’s syndrome,” *J. Intellect. Disabil. Res.*, vol. 37, no. 4, pp. 389–403, Jun. 1993.
- [38] L. L. Hardy, L. King, L. Farrell, R. Macniven, and S. Howlett, “Fundamental movement skills among Australian preschool children,” vol. 13, pp. 503–508, 2010.
- [39] D. Cáliz, J. Gómez, X. Alamán, L. Martínez, R. Cáliz, C. Terán, and V. Peñafiel, ““ Usability testing guide to mobile applications focused in persons with Down syndrome ‘ USATESTDOWN ’: Evaluation.”
- [40] A. Abran, A. Khelifi, W. Suryn, and A. Seffah, “Usability meanings and interpretations in ISO standards,” *Softw. Qual. J.*, vol. 11, no. 4, pp. 325–338, 2003.
- [41] J. Marco, E. Cerezo, and S. Baldassarri, “Bringing tabletop technology to all: Evaluating a tangible farm game with kindergarten and special needs children,” *Pers. Ubiquitous Comput.*, vol. 17, no. 8, pp. 1577–1591, 2013.
- [42] M. Corredentora, “Ma. Corredentora,” 2017. [Online]. Available: <http://mariacorredentora.org/wpmcorg/>. [Accessed: 10-Jan-2017].
- [43] A. ORG, “Apadema,” 1971. [Online]. Available: <http://www.apdema.org/tag/madrid/>.
- [44] “Fundación PRODIS,” 2016. [Online]. Available: <http://www.fundacionprodis.org/>. [Accessed: 14-Jun-2016].
- [45] J. Gómez, X. Alamán, G. Montoro, C. Juan, and A. Plaza, “AmICog – mobile technologies to assist people with cognitive disabilities in the work place,” *Adv. Distrib. Comput. Artif. Intell. Jornual*, vol. 2, no. 1, pp. 9–17, 2011.
- [46] Á. Fernández-López, M. J. Rodríguez-Fórtiz, M. L. Rodríguez-Almendros, and M. J. Martínez-Segura, “Mobile learning technology based on iOS devices to support students with special education needs,” *Comput. Educ.*, vol. 61, pp. 77–90, 2013.
- [47] D. Cáliz and X. Alamán, “Usability Evaluation Method for Mobile Applications for the Elderly : A Methodological Proposal,” pp. 252–260, 2014.
- [48] D. Cáliz, X. Alamán, L. Martínez, R. Cáliz, and V. Peñafiel, “Examining the Usability of Touchscreen Gestures for Elderly People.”
- [49] S. Rus and A. Braun, “Money Handling Training - Applications for Persons with Down Syndrome.”

10 Anexos.

USATESTDOWN

1. Establish the tasks

- Take the participants with the same level of disability.
- The tasks should not have a high complex level. The difficult scale should be of 1 to 3, where 1 is easier, the task should take 1.
- The session should be done in 10 minute sessions for each person,
- Do not limit to the participants the time.

2. Write the instructions

- Use only the oral way to explain the task to the participants
- Ask the participant if he/she is a willing participant.
- Speak slowly and repeat the task
- Perform the session with the participant individually.

3. Define the test plan

- The demographic survey must have only general information not lasts names.
- Apply the User satisfaction survey only with 3 answers . USATESRDOWN tets is recommended.
- Don't use documents with text to the participants. If you need to give documents it is better with graphics or don't not use documents
- Write short questions only with 3 answers
- Write questions focused on the applications and try to make understand the participants how important is to say the truth
- Participation of the expert tutors is required
- The sessions will be facilitated and observed by only one facilitator to one participant
- Persons need to be encouraged to participate and facilitator should stress the value of the child's input and show appreciation and gratitude.
- The educators of people with DS should be present at the moment of perform the session.
- Give the participant the device for long time to execute the game (or task) task at home. Application should have a data log to record.
- Apply at least 2 evaluation methods
- Establish objective metrics with a completion time for the task, error rate, etc.
- The participants should have pre training about the application
- Establish subjective metrics such as success, frustration, satisfaction, etc.
- Schedule a break halfway

- Record videos about the participant sessions but without faces.
- Hold a meeting prior to executing the test
- Observation method needs facilitator to record all children action, behaviours and facial expressions while observing children playing the game.
- The evaluator should evaluate every participant on an individual way.
- Do not use the technique of “thinking out loud”

4. Run the pilot test

- Record on video the interaction of the person with the mobile device
- Ask for parents’ permission if it is necessary to film faces,
- To the pilot test is necessary only one participant
- The tutors or professors should be present on the evaluation session.
- The facilitator will sit next to the child during the session to take note and solve participant questions
- Two questionnaires need to be assessed, they are proposed for USATESTDOWN
- Post task questionnaire will be conducted right after each test session with the help from the Tutor or Parent
- Take note of the times when the participant asks for help.
- Take a pause
- Use simple words when directing the participant
- Speak slowly and sometimes is necessary repeat the same idea

5. Refine the test plan after analysing the results of the pilot tests.

- Validate the tasks what the evaluator would you like to change with the Tutors
- Write the changes will be done and explain the reason why it was made.
- It is required a formal document signed by the participants tutor to make changes.

6. Recruit participants

- It is necessary that evaluators determine the cognitive disabilities level of the participants that they need, only the age is not a parameter to be on count.
- There should get at least 10 participants to the evaluation
- Send clear information about the session objective to the Technology Department, if they are interested on the research and they could find an importation to the centre they would contact with the Directives explaining the advantages they could have.
- Evaluator and Tutor should have a meeting to define the participant profile and tutor should recommend the participants profile to make the task that the evaluator proposed.
- The facilitator should ask the participant if he/she wants to collaborate
- Don’t push them to finish fast.
- Place should be the same where the participants work

7. Run the test session

- Do not complete the final test on the same day as the pilot testing
- Record everything the participant make and all his body behaviour
- Make a pause during the test
- Use simple language to speak with the participants
- Take note of the times when the participant asks for help.
- Speak slowly and repeat some times
- The participant who participated on the pilot test can participate on the real test
- Take not of the all participant behaviour
- Apply USATESTDOWN questionnaires' proposed
- Solve all of the questions that the user has during the process,
- Apply the questionnaires' USATESTDOWN propose as soon as possible.
- Ask the participant if he/she wants to collaborate
- Don't push the participant to answer a question

8. Analyse the collected information

- Compare the qualitative with qualitative dates.
- Take specific attention on applications data logs to make Statistics, it will show the real result
- The collected data by observing when the persons performing the task scenarios should be here
- The index indicators should be clear explained

9. Report results to the development team or management.

- All people involve in the test session should be on the presentation and give their own point of view and comments
- The document to send to the Design Develop department all the data logs information should be included.
- Make a short and easy to understand inform to the Directives people on the centre, they use to use the information to improve the education way
- Save the participants information on a confidential way after the test session.
- Include the old reports if it there would be from another interaction and make a comparison.

QUESTIONNAIRES'

SUS QUESTIONNAIRE	YES	MAYBE	NO
1. Would you use this application to work?			
2. Do you think the application is difficult to use?			
3. Do you think the application is easy to use?			
4. Do you think help is necessary in order to use the application?			
5. Do you think the application would be useful for your work colleagues?			

SUS Questionnaire for persons with Down syndrome

Open questions

1. What did you like mostly?.....
2. What did you dislike?.....
3. What did you find easy to do
4. What did you find so difficult to do ?.....

ID	Name	Age	Do you have a smart cell phone ?	Did you use computers ? How oft?
1				
2				
3				

Demographic Questionnaire example

		SUCCESS	SATISFACTION	FRUSTRATION	SUCCESS	SATISFACTION	FRUSTRATION
GENDER	USER ID	TASK 1			TASK N....		
F	3	0	0	1	1	1	3
F	4	1	1	1	1	1	2
M	6	1	1	1	1	1	3
F	10	0		1	1	1	2

Formulary Example to Tutor

User id	User Name	Age	Gender	Task Sequence No.	Evaluator	Do you have a smart cellphone ?	Ratio of tasks completed correctly /incorrectly (VIDEOS)	Number of times the user expresses satisfaction (VIDEOS)	Number of times the user expresses frustration
3	RR	28	MASCULINO	1	Doris	SI			
3	RC	28	MASCULINO	1	Doris	SI			
3	PP	28	MASCULINO	1	Doris	SI			

Examples of Metrics to measure the participant satisfaction level