## Summative and formative evaluations results

<table>
<thead>
<tr>
<th>Work package</th>
<th>Evaluation (WP7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>Task 7.2</td>
</tr>
<tr>
<td>Dissemination Level</td>
<td>Public</td>
</tr>
<tr>
<td>Publishing date</td>
<td>Contractual M30</td>
</tr>
<tr>
<td>Deliverable</td>
<td>D7.2</td>
</tr>
<tr>
<td>WP / Task responsible</td>
<td>UGOT</td>
</tr>
<tr>
<td>Contact person</td>
<td>Wolmet Barendregt, UGOT</td>
</tr>
<tr>
<td>Contributors</td>
<td>UGOT, JacobsUni, HWU, INESC-ID, UOB</td>
</tr>
<tr>
<td>Short abstract</td>
<td>This deliverable describes the formative and summative measures that will be used to evaluate the robotic tutor developed in EMOTE. It also describes the evaluation studies to be performed in the final year, as well as the results of the first formative evaluations for the two scenarios.</td>
</tr>
<tr>
<td>Keywords</td>
<td>Prototype evaluations</td>
</tr>
<tr>
<td>Documents</td>
<td></td>
</tr>
</tbody>
</table>
Executive Summary

This deliverable describes the formative and summative measures that will be used to evaluate the robotic tutor developed in EMOTE based on the evaluation strategy described in D7.1. It also describes the evaluation studies to be performed in the final year, as well as the results of the first formative evaluations for the two scenarios. Due to scheduling issues, these formative evaluations that should have taken place during this reporting period, have been of limited depth and breadth.
Contents

Executive Summary .................................................................................................................................................. 2

1 Introduction ...................................................................................................................................................... 5

2 Ethical aspects of participation ..................................................................................................................... 6
   2.1 Attachment to the robot ............................................................................................................................ 6
   2.2 Deception about the robot’s abilities ........................................................................................................ 6
   2.3 Authority and Autonomy .......................................................................................................................... 6
   2.4 Protocol for introduction .......................................................................................................................... 7

3 Formative and summative evaluations ........................................................................................................... 8

4 Summative measures and formative instruments ............................................................................................ 10
   4.1 ALMERE ............................................................................................................................................... 10
   4.2 NARS.................................................................................................................................................... 11
   4.3 Robot’s empathy .................................................................................................................................... 12
      4.3.1 Map reading scenario ...................................................................................................................... 12
      4.3.2 Sustainable development scenario ................................................................................................. 13
   4.4 Godspeed questionnaire .......................................................................................................................... 13
   4.5 Perceived learning ................................................................................................................................... 15
      4.5.1 Map reading system and Nao/Emys ................................................................................................. 15
      4.5.2 Enercities and Nao/Emys ............................................................................................................... 15
   4.6 Longterm User eXperience ...................................................................................................................... 16
      4.6.1 Usability of the application .............................................................................................................. 16
      4.6.2 Usefulness of the robotic tutor ........................................................................................................ 16
   4.7 Map reading skills .................................................................................................................................... 16
      4.7.1 Umeå .............................................................................................................................................. 16
      4.7.2 Hudiksvall ....................................................................................................................................... 18
      4.7.3 Results pilot test ............................................................................................................................... 20
   4.8 Perspective taking ...................................................................................................................................... 21
      4.8.1 Water supply ................................................................................................................................... 21
      4.8.2 Rubbish dump ................................................................................................................................... 22
      4.8.3 Results pilot test ............................................................................................................................... 22
   4.9 Sustainable development facts questionnaire ............................................................................................ 23
4.9.1 Results pilot test........................................................................................................................................25

5 Results of evaluations in Year 3 ..................................................................................................................26

5.1 Scenario 1 ..................................................................................................................................................26

5.1.1 Procedure and analysis ..........................................................................................................................26

5.1.2 Results ..................................................................................................................................................27

5.2 Scenario 2 ..................................................................................................................................................28

5.2.1 Procedure and analysis ..........................................................................................................................28

5.2.2 Results ..................................................................................................................................................29

6 References ....................................................................................................................................................30
1 Introduction

WP7 aims to evaluate the empathic tutors developed in EMOTE. It will (1) identify appropriate methods to evaluate the tutors, (2) iteratively evaluate the tutor’s prototypes both in laboratory and real classroom settings, and (3) provide input to WP6 for the development of the two showcases which will be evaluated in year 3 of the project. This document describes deliverable 7.2: Report of summative and formative evaluations of prototypes with students. Deliverable 7.2 contains the results of Task 7.2. Perform evaluation of initial tutor prototypes. This task aims to evaluate initial tutor prototypes with students. These evaluations can either be more summative, for example by letting students compare two or more prototypes on aspects like trust, likeability or perceived learning, but also more formative, for example, by observing students’ interactions with the tutors and how that may influence their learning processes.

In summary, the objectives of Task 7.2 are:

1. Determine when to perform the summative and formative evaluations with children (which stages of the prototypes, comparison of prototypes)
2. Adapt the summative evaluation methods to gather relevant measures from children
3. Determine most appropriate formative evaluation methods and develop coding schemes
4. Perform summative and formative evaluations with children and analyse the results in order to inform further design

In this deliverable we will describe the progress concerning each of these objectives in turn. However, we will first give the protocol to be followed for all contacts with children involving the robot. This protocol aims to ensure that, besides following the rules for informed consent, other ethical aspects regarding child-robot interaction are covered properly.
2 Ethical aspects of participation

The aim of EMOTE is to develop a robotic tutor that forms a social bond with lower-secondary students in order to promote learning in a personalised way. As Fridin (2014) describes for the use of assistive robots for pre-school children, this entails some ethical dilemmas, especially related to long-term interaction. These ethical dilemmas include attachment to the robot, deception about the robot’s abilities, and robot autonomy and authority. Of course, there are the ethical dimensions of informed consent, such as privacy. In this document we describe how the children should be informed about the robot before the start of the evaluation in order to deal with these ethical concerns without jeopardising the possibility to explore how social bonding can promote learning.

2.1 Attachment to the robot

In EMOTE the goal is to develop an empathic robotic tutor that students create a social bond with. Several studies have shown that people can form social and moral relationships with robots, e.g. (Kahn, Freier, Friedman, Severson, & Feldman, 2004). In a learning scenario, Saerbeck and colleagues (Saerbeck, Schut, Bartneck, & Janse, 2010) investigated the effects of social supportive behaviours of an iCat robot on children’s learning performance and motivation. The results indicate that simple manipulations in the robot’s supportiveness, while maintaining the same learning content, increased student’s motivation and scores on a language test. This means that if students do create social bond with the robot that supports learning, we have to inform them that this social bond will not be able to continue. Of course, students also form social bonds with humans that are only present in the school context for a shorter period, so when starting the evaluations we will explain how long exactly the robotic tutor will be present in the school and when it will be removed, similar to introducing a short-term teacher.

2.2 Deception about the robot’s abilities

In all our studies with children as well as teachers we have felt the need to be open about where the robot’s abilities come from in order not to deceive our participants. For the coming evaluations this means that we should allow the children to understand the robot’s workings by showing them the technology and answering any questions about it. However, since the empathic robotic tutor developed in EMOTE is still a prototype it is also tempting to voice ideas that the robot is stupid and unable to behave intelligently. Such explanations should be avoided since this may undermine the student’s ability to form a social bond with the robot, as is the purpose of the investigation.

2.3 Authority and Autonomy

The robot’s authority is another sensitive issue for the use of a robotic tutor. Besides forming a social bond with the child, the role of the robot is also to give the students directions requiring some measure of authority based on expertise. According to Murphy and Woods (2009) an imbalance between user autonomy and robot authority may create an ethical dilemma. In our studies with children (see e.g. Deliverable 7.1) we have found that children are very aware of this balance when they state that a robot in school should not be autonomous to such an extent that it is responsible for grading or for disciplining or will replace teachers. As Bryson (2009) suggests, the best way to get full utility from robots and to avoid any moral hazards is to consider robots as slaves. Thus, during our introductions we will explain that while the robot is trying to help them to accomplish learning
tasks in a pleasant manner, it will not be responsible for grading and does not have the authority to keep them engaged in the task if they want to drop out. The researcher is responsible for the robot and the teacher is responsible for all aspects related to the school context and the students’ education.

2.4 Protocol for introduction

Below follows the protocol to be used for the introduction of the child participants. This protocol is to be followed, but without reading it to the children.

Hi [name of the child/group],

I am [name] and I come from [university name]. Your school has agreed to help us evaluate the robotic tutor that we are developing in the EMOTE project. This robot will try to understand you and help you with tasks surrounding Geography and Sustainability. It will stay in the school for [number of weeks] weeks. The robot uses several advanced ways of trying to understand how to help you in the best way, which I will show you in a short video/in these pictures.

(show video or pictures of the set-up, allow children to ask questions)

Although you will be asked to work with this robot individually or in smalls group during the school days, the robot will not be able to force you to work with it, and your teacher is responsible for your grading and planning. The robot is programmed to help you with topics related to Geography and Sustainability and it will not do anything else by itself. If there are any technical problems you can always talk to me and I will try to fix it. In order for the robot to work it needs to record you, and we also want this data in order to improve out robot. As we also explained in the consent form which was sent to your parents, these recordings are only kept for research purposes. If you have any questions you can ask me now, or any time later.

I hope you will have fun with our robot!
3 Formative and summative evaluations

In D7.1 all evaluation studies to be performed during the final year of EMOTE were already described in general terms. These studies, while still working with showcases, are of a more summative nature although they all can have formative elements, e.g. to understand better how next generations of an empathic robotic tutor could be designed. Table 1 repeats these studies, but also indicates the exact summative measures which will be described in more detail in Chapter 3. Furthermore, the table mentions two measurement evaluation studies that will be performed by Jacobs University based on the data gathered from the children by the other partners.

Table 1 Summative evaluation studies

<table>
<thead>
<tr>
<th>Study description /conditions</th>
<th>Partners involved</th>
<th>Scenario(s) and conditions + participants or purpose of the study</th>
<th>Summative measures</th>
</tr>
</thead>
</table>
| C1: Physical Emys C2: Physical Nao | HWU | • Collaborative sustainable development  
• 20 students per condition  
• Small scale, non-school, university students | Sustainability facts (pre-post)  
Sustainability perspectives (pre-post)  
Perceived learning (post)  
ALMERE (post)  
NARS (pre-post)  
Godspeed (post)  
Empathy (post)  
Emergent Dialogue (during) |
| Migration study | HWU | • Individual map reading  
• 15 pairs of children in two groups | ALMERE (post)  
Map-reading (pre-post)  
Perceived learning (post)  
NARS (pre-post)  
Godspeed (post)  
Empathy (post) |
| C1: two children and the empathic robot  
C2: three children, no robot  
C3: two children and a non-empathic version of the robot | INESC | • Collaborative sustainable development  
• The first two conditions (C1 and C2) are used to determine whether there is difference in the learning criteria  
• The second and third condition (C2 and C3) are used to determine whether there is an additional value from the low-level empathic behaviours from the robot, such as looking at a player, mimicking behaviours etc. | Sustainability facts (pre-post)  
Sustainability perspectives (pre-post)  
Perceived learning (post)  
ALMERE (post)  
NARS (pre-post, but not for C2)  
Godspeed (post)  
Empathy (post)  
Emergent Dialogue (during) |
| C1: An empathic physical Nao robot  
C2: A non-empathic physical Nao robot | UOB & INESC | • Individual map reading  
• Between subjects, 40 children at UoB and 40 at INESC | Map-reading (pre-post)  
Perceived learning (post)  
ALMERE (post)  
NARS (pre-post)  
Godspeed (post)  
Empathy (post) |
| Longer term studies of the empathic robotic tutor using ethnographic and design-based research methods in different countries | UGOT & INESC | • Individual map reading and Collaborative sustainable development (UGOT)  
• Collaborative sustainable development (INESC)  
• Children in schools | Map-reading (pre-post)  
Sustainability facts (pre-post)  
Sustainability perspectives (pre-post)  
Perceived learning (post)  
ALMERE (post)  
NARS (pre-post)  
Godspeed (post)  
Empathy (post)  
MemoLine (post)  
Emergent Dialogue (during) |
| Comparison of the | Jacobs | • Individual learning task | Details to be discussed |
data from the portable (sensor)- devices and video recordings with standard lab equipment

### Fine-grained analysis of child-robot interaction

<table>
<thead>
<tr>
<th>Partners involved</th>
<th>Setup and participants</th>
<th>Formative methods</th>
<th>Evt. Summative measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>INESC</td>
<td>• Collaborative sustainable development scenario, Autonomous robot.</td>
<td>Observation with Thinking Aloud Post-task Interview about the game and the questionnaires</td>
<td>ALMERE Godspeed Empathy NARS</td>
</tr>
<tr>
<td>UGOT</td>
<td>• Individual map reading Scenario (2 trails), Autonomous robot.</td>
<td>Observation Post-task Interview</td>
<td></td>
</tr>
</tbody>
</table>

Throughout the project evaluations of a more formative nature have been performed, first on paper prototypes, thereafter using a WoZ methodology, and finally with the almost completely functional robot and applications. These evaluations have already been described in previous deliverables, such as the WoZ study and the mock-up study. During the final year of the project some additional evaluations have been performed with more full-fledged versions of the scenarios and the robotic tutor. However, since the final evaluations had to start before the summer holidays, the formative evaluations in the final year were of a rather limited nature. These evaluations are given in Table 2 and the results will be presented in chapter 5.

**Table 2 Formative evaluations performed during the final year**
4 Summative measures and formative instruments

This chapter describes all the (summative) measures and (formative) instruments to be used during the early and later evaluations of the tutor prototypes. All instruments will be translated to the children’s native tongue. However, for convenience, they are here given in English. Since the questionnaires will be used with different embodiments (Nao/Emys) this is indicated as an option. Each team will choose the wording to what is appropriate for their setup.

4.1 ALMERE

According to Heerink (2010) “[t]he Almere model is a technology acceptance model: it can be used to predict and explain usage of a system by observing the influences on the Intention to Use this system. This Intention to Use predicts actual usage of the system – for some systems also Facilitating Conditions and Social Influence are predictive influences on usage.” However, since children in school are not always allowed to determine their use of technology during school days, we only focus on the actual acceptance which might lead to a willingness to use but without necessarily being able to make decisions about future use.

Heerink, Kröse, Evers, and Wielinga (2010) used a five point scale in their experiments (totally agree – agree – neutral – do not agree – totally do not agree), which we will also apply. Below are the questions used, however, since the Almere model was originally developed for elderly we made some slight adaptations to fit children’s language skills as well as the context in which they are using the robot (e.g. teachers instead of staff). We also referred to the robot as Nao or Emys, since the intention is for children to form a social bond with it. These questions are supposed to be given in random order.

ANXIETY

1. If I should use Nao/Emys, I would be afraid to make mistakes with it
2. If I should use Nao/Emys, I would be afraid to break something
3. I find Nao/Emys scary
4. I find Nao/Emys intimidating

ATTITUDE

5. I think it’s a good idea to use Nao/Emys
6. Nao/Emys would make my schoolwork more interesting
7. It’s good to make use of Nao/Emys

FACILITATING CONDITIONS

8. I have everything I need to make good use of Nao/Emys
9. I know enough of Nao/Emys to make good use of it

PERCEIVED ADAPTIVITY

13. I think Nao/Emys can change itself to what I need
14. I think Nao/Emys will only do what I need at that particular moment
15. I think Nao/Emys will help me when I consider it to be necessary
PERCEIVED ENJOYMENT
16. I enjoy Nao/Emys talking to me
17. I enjoy doing things with Nao/Emys
18. I find Nao/Emys enjoyable
19. I find Nao/Emys fascinating

PERCEIVED EASE OF USE
20. I find Nao/Emys easy to use
21. I think I can use Nao/Emys without any help
22. I think I can use Nao/Emys when there is someone around to help
23. I think I can use Nao/Emys when I have good instructions

PERCEIVED SOCIABILITY
24. I consider Nao/Emys a nice partner to talk to
25. I find Nao/Emys nice to interact with
26. I feel Nao/Emys understands me
27. I think Nao/Emys is nice

PERCEIVED USEFULNESS
28. I think Nao/Emys is useful to me
29. It would be convenient for me to work with Nao/Emys in school
30. I think Nao/Emys can help me with many things in school

SOCIAL INFLUENCE
31. I think the teachers would like me working with Nao/Emys
32. I think many people would like me working with Nao/Emys

SOCIAL PRESENCE
33. When interacting with Nao/Emys I felt like I’m talking to a real person
34. It sometimes felt as if Nao/Emys was really looking at me
35. I can imagine Nao/Emys to be a living creature
36. I often think Nao/Emys is not a real person
37. Sometimes Nao/Emys seems to have real feelings

Trust
38. I would trust Nao/Emys if he gave me advice.
39. I would follow the advice Nao/Emys gives me.

4.2 NARS
To measure children’s attitudes towards robots in general, both before and after their experiences with the empathic robotic tutor developed in EMOTE we use a slightly adapted version of the NARS questionnaire (Nomura, Suzuki, Kanda, & Kato, 2006). The questionnaire is adapted to fit in the school context and contains the following questions:
<table>
<thead>
<tr>
<th>Nr</th>
<th>Question</th>
<th>Sub-scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I would feel uneasy if robots really had emotions.</td>
<td>S2</td>
</tr>
<tr>
<td>2</td>
<td>Something bad might happen if robots developed into living beings.</td>
<td>S2</td>
</tr>
<tr>
<td>3</td>
<td>I would feel relaxed talking with robots.*</td>
<td>S3</td>
</tr>
<tr>
<td>4</td>
<td>I would feel uneasy if I had to do my school work with help of a robot.</td>
<td>S1</td>
</tr>
<tr>
<td>5</td>
<td>If robots had emotions I would be able to make friends with them.*</td>
<td>S3</td>
</tr>
<tr>
<td>6</td>
<td>I feel good being with robots that have emotions.*</td>
<td>S3</td>
</tr>
<tr>
<td>7</td>
<td>The word “robot” means nothing to me.</td>
<td>S1</td>
</tr>
<tr>
<td>8</td>
<td>I would feel nervous using a robot in front of other people.</td>
<td>S1</td>
</tr>
<tr>
<td>9</td>
<td>I would hate the idea that robots or intelligent computers were making judgments about things</td>
<td>S1</td>
</tr>
<tr>
<td>10</td>
<td>I would feel very nervous just standing in front of a robot.</td>
<td>S1</td>
</tr>
<tr>
<td>12</td>
<td>I feel that if I depend on robots too much, something bad might happen.</td>
<td>S2</td>
</tr>
<tr>
<td>12</td>
<td>I would feel paranoid talking with a robot.</td>
<td>S1</td>
</tr>
<tr>
<td>13</td>
<td>I am concerned that robots would be a bad influence on younger children.</td>
<td>S2</td>
</tr>
<tr>
<td>14</td>
<td>I feel that in the future society will be dominated by robots.</td>
<td>S2</td>
</tr>
</tbody>
</table>

This version of the questionnaire was already successfully used during one of the previous studies in EMOTE to study perceptions before and after interacting with a robotic tutor (Serholt, Basedow, Barendregt, & Obaid, 2014).

### 4.3 Robot’s empathy

To measure children’s perceptions of the robot’s empathic capacities we will use the questionnaire by Davis (1983) in an adapted form. Instead of asking about children’s perceptions of their own empathic capacities they are asked to gauge the robot’s empathy. The questions vary slightly between the two different scenarios since in the map reading scenario the children do not have any experience with the robot’s behavior towards others.

#### 4.3.1 Map reading scenario

<table>
<thead>
<tr>
<th>Nr</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nao/Emys can have tender and concerned feelings for people less fortunate than himself.</td>
</tr>
<tr>
<td>2</td>
<td>Sometimes Nao/Emys found it difficult to see things from my point of view.</td>
</tr>
<tr>
<td>3</td>
<td>Sometimes Nao/Emys did not feel sorry for me when I was having problems.</td>
</tr>
<tr>
<td>4</td>
<td>Nao/Emys tries to look at all sides of an issue before he makes a decision.</td>
</tr>
<tr>
<td>5</td>
<td>If Nao/Emys would see someone being bothered or hurt, he would probably feel protective towards them.</td>
</tr>
<tr>
<td>6</td>
<td>Nao/Emys sometimes tried to understand me better by imagining how things look from my perspective.</td>
</tr>
<tr>
<td>7</td>
<td>Nao/Emys is not disturbed when I or someone else am upset.</td>
</tr>
<tr>
<td>8</td>
<td>When Nao/Emys is sure he is right about something, he doesn’t waste much time listening to my arguments or point of view.</td>
</tr>
<tr>
<td>9</td>
<td>If Nao/Emys would see someone being treated unfairly, he wouldn’t feel much pity for them.</td>
</tr>
<tr>
<td>10</td>
<td>Nao/Emys is often quite touched by things that he sees happening.</td>
</tr>
<tr>
<td>11</td>
<td>Nao/Emys believes that there are two sides to every question and tries to look at them both.</td>
</tr>
<tr>
<td>12</td>
<td>I would describe Nao/Emys as a pretty soft-hearted robot.</td>
</tr>
<tr>
<td>13</td>
<td>If Nao/Emys were upset, he would try to “put himself in my shoes” for a while to understand the situation.</td>
</tr>
<tr>
<td>14</td>
<td>Before telling me I have done something wrong, Nao/Emys tries to imagine how he would feel if he was in my place.</td>
</tr>
</tbody>
</table>
4.3.2 Sustainable development scenario

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nao/Emys often had tender, concerned feelings for people less fortunate than himself.</td>
</tr>
<tr>
<td>2</td>
<td>Sometimes Nao/Emys found it difficult to see things from someone else’s point of view.</td>
</tr>
<tr>
<td>3</td>
<td>Sometimes Nao/Emys did not feel sorry for other people when they were having problems.</td>
</tr>
<tr>
<td>4</td>
<td>Nao/Emys tried to look at everybody’s side of a disagreement before making a decision.</td>
</tr>
<tr>
<td>5</td>
<td>When Nao/Emys saw someone being taken advantage of, he felt kind of protective towards them</td>
</tr>
<tr>
<td>6</td>
<td>Nao/Emys sometimes tried to understand me better by imagining how things look from my perspective</td>
</tr>
<tr>
<td>7</td>
<td>Other people's misfortunes do not usually disturb Nao/Emys a great deal</td>
</tr>
<tr>
<td>8</td>
<td>If Nao/Emys was sure about something, he didn’t waste much time listening to other people’s arguments.</td>
</tr>
<tr>
<td>9</td>
<td>When Nao/Emys saw someone being treated unfairly, he didn’t feel very much pity for them</td>
</tr>
<tr>
<td>10</td>
<td>Nao/Emys is often quite touched by things that he sees happening</td>
</tr>
<tr>
<td>11</td>
<td>Nao/Emys believes that there are two sides to every question and tries to look at them both.</td>
</tr>
<tr>
<td>12</td>
<td>I would describe Nao/Emys as a pretty soft-hearted robot.</td>
</tr>
<tr>
<td>13</td>
<td>When Nao/Emys is upset at someone, he usually tries to “put itself in the others’ shoes” for a while.</td>
</tr>
<tr>
<td>14</td>
<td>Before criticizing somebody, Nao/Emys tries to imagine how it would feel if it was in their place.</td>
</tr>
</tbody>
</table>

Answers to each question will go from “this does not describe Nao/Emys at all” to “this describes Nao/Emys very well” in a Likert scale from 1-5 points.

4.4 Godspeed questionnaire

The Godspeed questionnaires defined by Bartneck, Kulic, and Croft (2009) will be used to gauge children’s impressions of the robot, going beyond the ones already covered in the ALMERE questionnaire.

**Anthropomorphism**

Please rate your impression of Nao/Emys on these scales:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Natural</td>
</tr>
<tr>
<td>Machinelike</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Unconscious</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Artificial</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Moving rigidly</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Animacy**
Please rate your impression of the robot on these scales:

**Dead**  
1 2 3 4 5 **Alive**

**Stagnant**  
1 2 3 4 5 **Lively**

**Mechanical**  
1 2 3 4 5 **Organic**

**Artificial**  
1 2 3 4 5 **Lifelike**

**Inert**  
1 2 3 4 5 **Interactive**

**Apathetic**  
1 2 3 4 5 **Responsive**

**Likeability**

Please rate your impression of the robot on these scales:

**Dislike**  
1 2 3 4 5 **Like**

**Unfriendly**  
1 2 3 4 5 **Friendly**

**Unkind**  
1 2 3 4 5 **Kind**

**Unpleasant**  
1 2 3 4 5 **Pleasant**

**Awful**  
1 2 3 4 5 **Nice**

**Perceived intelligence**

Please rate your impression of the robot on these scales:

**Incompetent**  
1 2 3 4 5 **Competent**

**Ignorant**  
1 2 3 4 5 **Knowledgeable**

**Irresponsible**  
1 2 3 4 5 **Responsible**

**Unintelligent**  
1 2 3 4 5 **Intelligent**

**Foolish**  
1 2 3 4 5 **Sensible**

**Perceived safety**

Please rate how you felt on these scales at the beginning:

**Anxious**  
1 2 3 4 5 **Relaxed**

**Agitated**  
1 2 3 4 5 **Calm**

**Quiescent**  
1 2 3 4 5 **Surprised**
Please rate how you felt on these scales towards the end:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxious</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Relaxed</td>
</tr>
<tr>
<td>Agitated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Calm</td>
</tr>
<tr>
<td>Quiescent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Surprised</td>
</tr>
</tbody>
</table>

Based on the work by Kennedy, Baxter, and Belpaeme (2014) we added the following final question that summarizes the experienced relationship with the robot:

Please indicate what Nao/Emys was for you (put a cross in the table under the word that fits best):

<table>
<thead>
<tr>
<th>Brother or sister</th>
<th>Classmate</th>
<th>Stranger</th>
<th>Relative (e.g. cousin or aunt)</th>
<th>Friend</th>
<th>Parent</th>
<th>Teacher</th>
<th>Neighbour</th>
</tr>
</thead>
</table>

4.5 Perceived learning

For children’s perceptions of learning we will use a questionnaire developed by Robertson and Cross (2004) for the evaluation of an intelligent tutoring system with children between 10 and 12 years old. Although Robertson and Cross used a visual analogue scale (VAS) for this questionnaire, research has shown that Likert-scales are preferred over VAS (van Laerhoven, van der Zaag-Loonen, & Derkx, 2004), so we will use a 5-point Likert-scale instead, ranging from 1= Completely disagree to 5 = Completely agree.

4.5.1 Map reading system and Nao/Emys

1. I enjoyed working with the map reading system and Nao/Emys
2. I think the map reading system and Nao/Emys made my map reading skills worse
3. I think I would like to use the map reading system and Nao/Emys again
4. Using the map reading system and Nao/Emys helped me to become better at map reading
5. I found using the map reading system and Nao/Emys confusing
6. The system and Nao/Emys make map reading easier
7. I think I need someone to help me use the map reading system and Nao/Emys
8. I think the system’s and Nao’s/Emys’ advice on map reading was useful
9. A teacher is more helpful than the map reading system and Nao/Emys
10. The map reading system with Nao/Emys is boring

4.5.2 Enercities and Nao/Emys

1. I enjoyed working with Enercities and Nao/Emys
2. I think Enercities and Nao/Emys made my knowledge on sustainability worse
3. I think I would like to use Enercities and Nao/Emys again
4. Using Enercities and Nao/Emys helped me to think more about sustainability
5. I found using Enercities and Nao/Emys confusing
6. Enercities and Nao/Emys make thinking about sustainability easier
7. I think I need someone to help me use Enercities and Nao/Emys
8. I think Enercities’ and Nao’s/Emyss advice about building a sustainable city was useful
9. A teacher is more helpful to understand how to build a sustainable city than Enercities and Nao/Emys
10. Enercities with Nao/Emys is boring

4.6 Longterm User Experience
To help children to reflect upon the experience of working with the robot over a longer period of time the MemoLine (Vissers, de Bot, & Zaman, 2013) instrument will be applied. Instead of focusing merely on the usability of the application we will also ask explicitly about the usefulness of the robotic tutor.

4.6.1 Usability of the application

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>I understood exactly what I was supposed to do and how</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I did not understand what I was supposed to do or how</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.6.2 Usefulness of the robotic tutor

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>The robot helped me very well</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot did not help me very well</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.7 Map reading skills
To measure children’s map reading skills a pre-and post-test will be used. Two map reading tests have been piloted with 48 children of grade 4 and 5 in order to determine whether their difficulty levels are equal and whether they are not too easy for the target group. Based on these results the final pre- and post-tests will be created.

4.7.1 Umeå
There are two maps: 1. a map of the whole of Umeå, 2. a part of Umeå (detail map). In the questions below you will use both, but it is always indicated which one or which ones.
1. Whole map: There is a university towards the right of the map (universitetet). What is the name of the residential area that lies almost straight to the north of the university? ________________

2. Whole map: Find Röbäcks kapell (Röbacks chapel). Where do you get if you go 1.1 kilometer to the southeast? Write a 1 on the map where you get.

3. Whole map: Where do you get if you go 750 meter straight to the south-southwest from the central station? Write a 2 on the map where you get.

4. Whole map + Detail map: On the big map there is an island in the river to the north of the residential area called Böleång. Now check the detail map: What is the name of the part of the island that lies most to the west? ____________________________

5. Detail map: There is a power line over Lundåkern. How long is this power line?

6. Whole map: The city of Umeå considers building a new rubbish dump. There are six possible places that are indicated with numbers. Find the correct place based on the following clues:
   a. First, find the word “Start” on the map and check where the arrow points at. The rubbish dump should not be built to the west of this point.
   b. The rubbish dump should not be built in a forest area
   c. The rubbish dump may not be placed further than 2 kilometer from the start point

   Write the number of the place where the rubbish dump should be built: ____________________________
4.7.2 Hudiksvall

1. Whole map: There is a hospital (sjukhus) in the middle of the map. What is the name of the nearest lake that lies almost straight to the east of the hospital?

2. Whole map: Where do you get if you go 0.75 kilometer to the north-west from the central station? Write a 1 on the map where you get.

3. Whole map: Find Köpmanberget (there is an arrow pointing it out). Where do you get if you go 1300 meter to the northnortheast? Write a 2 on the map where you get.

4. Whole map + Detail map: On the whole map you will see two green islands to the east of Hudiksvall. Check the detail map: What is the name of the island that lies most to the west?

5. Detail map: There is a bathing area (Badplats) close to Maln. What is the distance from Håstaholmen to this bathing area by boat?

6. Whole map: The city of Hudiksvall considers building a new water supply. There are six possible places that are indicated with numbers. Find the correct place based on the following clues:
   a. First, find the word “Start” on the whole map and check where the arrow is pointing. The water supply may not be built to the west of this point.
   b. The water supply must be built in a forested area
   c. The water supply must lie around 2 kilometer from the start point

Write the number of the place where the water supply should be built: ____________________
4.7.3 Results pilot test
Each question was given one point if answered correctly. The results of the pilot test were as follows:

Table 3 Average results per question for each map

<table>
<thead>
<tr>
<th>Map</th>
<th>Map1</th>
<th>Map2</th>
<th>Map3</th>
<th>Map4</th>
<th>Map5</th>
<th>Map6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umeå</td>
<td>.78</td>
<td>.30</td>
<td>.04</td>
<td>.43</td>
<td>.17</td>
<td>.39</td>
</tr>
<tr>
<td>N</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.422</td>
<td>.470</td>
<td>.209</td>
<td>.507</td>
<td>.388</td>
<td>.499</td>
</tr>
<tr>
<td>Hudiksvall</td>
<td>.96</td>
<td>.00</td>
<td>.20</td>
<td>.36</td>
<td>.32</td>
<td>.36</td>
</tr>
<tr>
<td>N</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.200</td>
<td>.000</td>
<td>.408</td>
<td>.490</td>
<td>.476</td>
<td>.490</td>
</tr>
<tr>
<td>Total</td>
<td>.88</td>
<td>.15</td>
<td>.13</td>
<td>.40</td>
<td>.25</td>
<td>.37</td>
</tr>
<tr>
<td>N</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>.334</td>
<td>.357</td>
<td>.334</td>
<td>.494</td>
<td>.438</td>
<td>.489</td>
</tr>
</tbody>
</table>

This table shows that the first question, using only a basic direction and no distances is relatively easy. The fourth question, which involves switching maps of different scales is slightly more difficult. Questions using distances and symbols are the most difficult.

Table 4 Overall results for each map

<table>
<thead>
<tr>
<th>Mapreading</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Umeå</td>
<td>23</td>
<td>2.13</td>
<td>1.424</td>
<td>.297</td>
</tr>
<tr>
<td>Hudiksvall</td>
<td>25</td>
<td>2.20</td>
<td>1.354</td>
<td>.271</td>
</tr>
</tbody>
</table>

There is no significant difference between the two map reading exercises.

For our summative evaluations we are going to use the same two maps but they will be made neutral and names of locations will be fictive. In order to give the children a sense of accomplishment, we will include the relatively easy kinds of questions as well.
4.8 Perspective taking

To test children’s ability to understand that there are always many different perspectives we created two different writing assignments as pre- and post-test. Both tests have been tested with 48 children between of grade 4 and 5 in order to determine whether their difficulty levels are equal and whether they are not too easy for the target group. The measurement taken is a judgement of how many different perspectives are mentioned in the argument for a certain solution. Three perspectives that children could at least mention are Environment, Economy, and Wellbeing (as the ones present in the game). The places are fictive allowing the children to reflect freely upon the alternatives without being constrained by reality.

4.8.1 Water supply

The number of inhabitants in the city Kingsville is increasing and has therefore a problem with its water supply. During some periods there is not enough water in the wells. There are several possible solutions to solve this problem, and the city would like to get your advice on which solution to choose:

1. Drill deeper wells to take more water from the ground.
2. Use the water from the sea and remove the salt.
3. Buy water from the city Queenshight which is about 20 kilometers away and which has more than enough water.
4. Make a dam in the river Fishing close to the city and build a new water supply.
5. Restrict the inhabitants’ use of water.
6. Reconstruct all public water systems in the city in order to recycle and clean the water so that it can be reused.

A simple map of the area is given below:

Describe what you would recommend the city and WHY. There is no single correct answer! If you think that there are alternative solutions or if you think you need more information to make a choice between several answers, you can also write this down, and say what kind of information you would need.
4.8.2 Rubbish dump
The rubbish dump of the city Wellhouse has become too small. There are several possible solutions to solve this problem, and the city would like to get your advice on which solution to choose:

1. Build a new rubbish dump close the old one.
2. Build a waste incinerator (where rubbish is burned) close to the city.
3. Transport all waste to the city Sparrow, about 50 kilometers away, where they do not have any such problems.
4. Give all inhabitants a waste container with different compartments to sort their rubbish for recycling. This way the amount of waste for the old rubbish dump can be decreased so that it is big enough.
5. Give all inhabitants a limit for how much waste they are allowed to have. If they have more they will have to pay.
6. Force all food shops in the city to refrain from using plastic bags and unnecessary packaging. All inhabitants must bring their own bags and containers.

A simple map of the area is given below:

Describe what you would recommend the city and WHY. There is no single correct answer! If you think that there are alternative solutions or if you think you need more information to make a choice between several answers, you can also write this down, and say what kind of information you would need.

4.8.3 Results pilot test
Each different perspective was given one point if mentioned. The results of the pilot test were as follows:

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of perspectives</td>
<td>Water</td>
<td>23</td>
<td>.70</td>
<td>.926</td>
</tr>
<tr>
<td></td>
<td>Waste</td>
<td>25</td>
<td>.52</td>
<td>.770</td>
</tr>
</tbody>
</table>

Many children found it very hard to give any reasons for their choice and just gave one of the numbers without any argument.
4.9 Sustainable development facts questionnaire

Based on the factual information available in the game a questionnaire was created. This questionnaire (see below) has been pilot tested with 48 children between of grade 4 and 5 in order to determine whether the difficulty level is appropriate and to be able to split the questionnaire in two comparable halves, based on the percentage of correct answers to each question. These two halves will be used as the pre- and post-test.

1. Which factors need to cooperate to create a sustainable city?

Place a cross in front of the correct answer (only one answer is correct)

1. _______ Economical and social factors
2. _______ Economical, social, and environmental factors
3. _______ Environmental and social factors
4. _______ Economical, environmental and technological factors

2. How is a city’s sustainability affected by energy generation?

Circle Right or Wrong for each of the propositions below:

Right/Wrong  Power plants have environmental costs because they affect the environment
Right/Wrong  Different energy sources all have the same influence on a city’s economy
Right/Wrong  Power plants cost too much and are therefore not part of a sustainable city

3. A small coal power plant:

Place a cross in front of the correct answer (only one answer is correct)

1. _______ Uses water power to create a stable energy stream
2. _______ Uses the sun and creates a lot of energy
3. _______ Produces electricity by burning coal. It is small, cheap, and influences the environment negatively
4. _______ Does not affect the air quality but has to store uranium

4. Circle Right or Wrong for each of the propositions below:

Right/Wrong  Policies only have a positive effect on the environment
Right/Wrong  Policies do not cost anything
Right/Wrong  It is smart to introduce policies when one is planning a city
Right/Wrong  Policies can affect a city’s economy negatively

5. Indicate which policies can help to decrease the use of non-renewable resources (such as oil).

Circle Right or Wrong for each of the propositions below:

Right/Wrong  CO2 taxes
Right/Wrong  Electric car grid
Right/Wrong  Energy education programs
Right/Wrong  Ecotourism program
Right/Wrong  Fund for sustainable technology

6. Order the following power plants according to their influence on the environment.
Use 1, 2, 3 and 4 (each number once). 1 is least influence, 4 most

________ Small coal power plant
________ Water power plants
________ Nuclear power plant
________ Windmills

7. Which investments can help to decrease the use of non-renewable resources, such as oil?
Place a cross in front of all correct answers (there can be more than one)

________ Recycling stations
________ Solar roofs
________ Better isolation
________ Metro station
________ Better uranium storage

8. Improving the economy by building e.g. commercial constructions influences:
Place a cross in front of the best answer (only one answer is correct)

1. ________ The economy improves but it is bad for the environment, and for the use of renewable and non-renewable forms of energy.
2. ________ The economy improves but the use of non-renewable energy increases.
3. ________ The economy improves but it is bad for the environment.
4. ________ The economy improves but it is bad for people’s wellbeing, the environment, and the use of non-renewable forms of energy.

9. Public services, such as hospitals, police, and fire departments can improve people’s wellbeing if: Place a cross in front of all correct answers (there can be more than one)

________ They are built close to residential areas so that many people have access to them.
________ They are constructed near forests to balance the distribution of constructions in a city.
These services have costs in terms of energy and oil, so they should not be part of a sustainable city.

1. **By changing some behaviors, citizens can directly influence the city sustainability.**

**Circle Right or Wrong for each of the behaviors below:**

- Right/Wrong Buy organic/biologically grown food
- Right/Wrong Use public transport
- Right/Wrong Eat vegetarian food

4.9.1 **Results pilot test**

On the questionnaire it was possible to score between 0 and 28 points. The results are as follows:

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>18</td>
<td>13.17</td>
<td>4.890</td>
<td>1.153</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>14.43</td>
<td>6.027</td>
<td>1.100</td>
</tr>
</tbody>
</table>

There was no significant difference between the children in the two grades.

The result for each question separately are as follows:

<table>
<thead>
<tr>
<th>Question</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SustFact1</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.52</td>
<td>.505</td>
</tr>
<tr>
<td>SustFact2a</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.60</td>
<td>.494</td>
</tr>
<tr>
<td>SustFact2b</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.46</td>
<td>.504</td>
</tr>
<tr>
<td>SustFact2c</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.65</td>
<td>.483</td>
</tr>
<tr>
<td>SustFact3</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.42</td>
<td>.498</td>
</tr>
<tr>
<td>SustFact4a</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.50</td>
<td>.505</td>
</tr>
<tr>
<td>SustFact4b</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.60</td>
<td>.494</td>
</tr>
<tr>
<td>SustFact4c</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.54</td>
<td>.504</td>
</tr>
<tr>
<td>SustFact4d</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.58</td>
<td>.498</td>
</tr>
<tr>
<td>SustFact5a</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.37</td>
<td>.489</td>
</tr>
<tr>
<td>SustFact5b</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.46</td>
<td>.504</td>
</tr>
<tr>
<td>SustFact5c</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.46</td>
<td>.504</td>
</tr>
<tr>
<td>SustFact5d</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.33</td>
<td>.476</td>
</tr>
<tr>
<td>SustFact5e</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.31</td>
<td>.468</td>
</tr>
<tr>
<td>SustFact6</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.15</td>
<td>.357</td>
</tr>
<tr>
<td>SustFact7a</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.63</td>
<td>.489</td>
</tr>
<tr>
<td>SustFact7b</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.62</td>
<td>.489</td>
</tr>
<tr>
<td>SustFact7c</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.23</td>
<td>.425</td>
</tr>
<tr>
<td>SustFact7d</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>.52</td>
<td>.505</td>
</tr>
</tbody>
</table>
Based on these results the pre-test and post-test will contain the following questions:

<table>
<thead>
<tr>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2a,b,c</td>
<td>4b, 4c</td>
</tr>
<tr>
<td>4a,d</td>
<td>5a,b,c,d,e</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>7a,b,c,d,e</td>
<td>10a,b,c</td>
</tr>
<tr>
<td><strong>12 items, average correct 6</strong></td>
<td><strong>12 items, average correct 5.8</strong></td>
</tr>
</tbody>
</table>

Question 9 has been removed because of the high percentage of correct answers. The post-test is slightly harder than the pretest, but not significantly so.

5 Results of evaluations in Year 3

During the whole project, several formative evaluations have been performed, using paper prototypes of the task, a Wizard of Oz (WoZ), and a WoZ with limited perception for the wizard. Coming closer to the evaluations of the robotic tutors in the classroom some formative evaluations have been performed with nearly finished versions of the two scenarios.

When performing these formative evaluations we consider the robot and the tabletop application as one integrated system, with which the children interact.

5.1 Scenario 1

An informal formative evaluation was conducted with an almost final version of the prototype of Scenario 1. For the longitudinal study aimed to be conducted in a Swedish primary school, approximately 18 trails are under development comprising 6 different cities in Sweden.

5.1.1 Procedure and analysis

The experimental material consisted of two different trails, of which the first was the trail aimed for session 1 during the long-term evaluation (Malmö_Medium) and the second trail was aimed for a later session (Arjeplog_Easy). As the names indicate, two different levels of difficulty were chosen in order to determine feasibility of different difficulty levels and how these could relate to children’s
variations in Zones of Proximal Development (ZPD) (Vygotsky, 1978 (Original manuscripts [ca. 1930-1934])). The most advanced level was excluded here due to participant’s age.

The participant was a male at the age of 6½. The experiment was carried out in the presence of a researcher with a teaching background, along with the participant’s younger sibling. As the participant was younger than the target group in EMOTE, extra scaffolding was provided by the researcher present in accordance with the strategy used by the empathic robot. In other words, the strategies commonly observed in teacher-student interactions (see D2.2) in order to facilitate progression in a map-reading task, which has later been adapted for the EMOTE robot, were practiced by the researcher in order to test the feasibility of the strategy. Any technical issues hindering the robot from carrying out the proper strategy were also noted.

In this version of the scenario the perception module was not tested, meaning that the robot was unable to pick up on potential socio-emotional signals expressed by the child. Therefore, the researcher had to intervene several times in order to assist the child in both understanding the robot and understanding the task.

The experiment was video-recorded and field notes were made. This data led to the following recommendations.

5.1.2 Results

5.1.2.1 Formulation of instructions
Based on feedback from project reviewers, it was decided to reduce the amount of speech generated by the robot in order to allow for a more self-directed activity for the child. However, during the evaluation the child struggled clearly to figure out what the task required, suggesting that this kind of task requires scaffolding at least at the beginning of the activity in order to provide clear guidelines for the child in terms of what is required during the task. This should occur at the onset of the activity.

There should furthermore be clear and more elaborative feedback (from the robot and/or the tabletop application) when something remarkable occurs during the task (e.g. a clue or treasure is found).

5.1.2.2 Pedagogical strategy
During the test sessions a steady decrease in attention towards the robot was observed, where the child’s attention shifted from the robot to the researcher once the child deemed the robot’s assistance unhelpful. Of course, this is partly due to the lack of the perception module. Therefore, the perception module should be tested along with the scenario in order to acquire a true representation of the pedagogical strategy. Although repetitions were not reacted upon by the child very much, the utterances generated should vary so as to increase the chance of providing appropriate assistance to the child.

In terms of the level of difficulty of the task, the child in the study required much assistance from the researcher present, which can be partly ascribed to the child’s age. Based on the teaching experience of the researcher, it was assessed that the level of difficulty would be feasible for the
target age group, as the level Easy was just outside the child’s Zone of Proximal Development. In other words, the easiest level was slightly too difficult for the child to accomplish even in the presence of a mentor (in this case the researcher). This means that the researcher sometimes had to provide the answers (e.g. in reading of the map key containing references to the map symbols, as the child was not yet at the stage of fluent reading).

5.1.2.3 Robot behaviours
During the session, sudden bodily movements of the robot alarmed the child, as were certain eye gaze behaviours developed in the EMOTE project. The child explained apprehensions surrounding this as a fear that the robot was about to break down. This can be related to the child’s previous experience with robots not having these behaviours. In other words, the child’s expectations were not met, and perhaps certain behaviours were associated with experience of other technologies breaking down. Nevertheless, previous research on children’s preconceptions and expectations of robots has suggested that these do not necessarily affect the interaction in the longer term (Belpaeme et al., 2012). Therefore, we do not advise any changes to these behaviours yet.

5.1.2.4 Verbal interactions
The importance of verbal interaction was prevalent for a child at this age. Indeed, it could be the case that the presence of the researcher triggered much verbal inquiry about the task from the child towards both the robot and the researcher. This, coupled with the fact that the task was too advanced for the child, could have directly resulted in this verbal inquiry. Although speech recognition will not be developed within the EMOTE project, it seems that children have a strong tendency to voice their thoughts at this age, which means that a recognition of keywords could be fruitful in order for a robot to acquire important input for the pedagogical strategy as a whole.

Take the following example: if the robot instructs, “Please go 300 meters North until you see a bus stop” and the child responds by saying, “North?! Which way was that?”, the robot can infer along with other inputs that the child may not know where “north” is. Although it can just as easily be the case that the child said, “Ok. North”, the robot can nonetheless respond “North is illustrated by N on the compass” and still appear socially and pedagogically competent. As we are simply dealing with possibilities in HRI due to limitations in technological advancement (Mubin, Stevens, Shahid, Mahmud, & Dong, 2013), the robot does not need to be certain in its perception of its surrounding world; rather, we are consistently dealing with probabilities.

5.2 Scenario 2

5.2.1 Procedure and analysis
The formative evaluation of the collaborative sustainable scenario (EMOTE scenario 2) was performed with an autonomous version of the empathic robotic tutor playing Enercities in two different sessions. Four children, two girls and two boys, (mean age = 13.75) participated in the formative evaluation of the scenario. The evaluation started with a Thinking Aloud Technique, a method that involves asking users to think out loud as they are performing the task. So, children were asked to say whatever they were looking at, thinking, and feeling at each moment. This is a useful method to determine users’ expectations and identify aspects of the system that are
confusing and in need of refinement (Nielsen, 1992). As this technique involves the capacity of the users to elaborate on their own thoughts and feelings without a filtered system, this becomes specially challenging to apply with children. It is therefore advised to involve the children in a training session (Markopoulos, Read, MacFarlane, & Hoysniemi, 2008). The activity children performed was about the “fox, goose, and bag of beans” classical puzzle, in which the story was the following:

*Once upon a time a farmer went to market and purchased a fox, a goose, and a bag of beans. On his way home, the farmer came to the bank of a river and rented a boat. But in crossing the river by boat, the farmer could carry only himself and a single one of his purchases - the fox, the goose, or the bag of the beans. If left together, the fox would eat the goose, or the goose would eat the beans. The farmer's challenge was to carry himself and his purchases to the far bank of the river, leaving each purchase intact. How did he do it?*

There are several answers to this puzzle, which constitutes an interesting and engaging task for children to practice the thinking-aloud technique. To solve this puzzle, children were ask to think aloud about the answer in a process that would familiarise them when exposed to the real Emote scenario.

After 10 minutes of trying to solve this puzzle, children moved to the learning setting where the robot and the Enercities game were present. They played the game for 20 minutes and were being prompted by a researcher who asked them to verbally express their thinking about the game and about the robot.

At the end of each session, children were asked by the researcher in a post-task interview to elaborate on some aspects that were picked up during the game. Overall, children liked to play with the robot and engage in the task wherever they could. This demonstrates that both the robot and the task were jointly working as a balanced educational setting. Nonetheless, there were some aspects that children pointed out and that required revision. These aspects will be discussed next.

5.2.2 Results

5.2.2.1 Formulation of instructions
At the beginning the robot explained the tutorial to children. As Enercities is a rather complex and a positively challenging game about sustainability that requires giving attention to a considerable amount of game features. Children expressed they felt sometimes lost during the process of acknowledging all that information and integrating it with the game play. To solve this issue, the tutorial’s dynamic was refined in the following way: The robot now performs an interactive tutorial, where it explains to children the rules of the game in the context where the features emerge and deserve being explained. So, for e.g., if the robot is constructing a building, it will explain to children the rules for how to perform constructions in the game.

5.2.2.2 Verbal interactions
Children indicated that they felt sometimes the robotic tutor talked for too long and too quickly. This was solved by tackling the longer verbal behaviours of the robot and splitting them into different
utterances, to provide intonation to each idea that is expressed by the robot. For the cases when the tutor was perceived as talking too quickly, the verbal behaviour of the robot was refined according to Nao’s Aldebaran text to speech report, and some conventions, like pauses during the speech, were integrated.

Children also felt that the robot sometimes lacked contingency behaviours, such as asking a question but not giving any feedback after the child’s answer. This happens due to the fact that EMOTE is not using any speech recognition technology as it is still too unreliable during interactions. To solve this issue, the robot now answers contingently to children after having asked a question which is made possible by a technical implementations described in D5.3. The tutor is now aware that it has asked a question, and therefore, in these situations, a contingency answer is needed, such as “I see your point.”

5.2.2.3 Interaction with the game

Regarding the Enercities game, children found it interesting, but sometimes they struggled with double clicks required during the game to, e.g. make a construction. This made children insecure about their performance, as they were not expecting having to perform a double click. Instead they used a single click. This unexpected interaction was removed and now a single click is required to interact with the game interface.

5.2.2.4 Evaluation of the questionnaires

After the think-aloud session and the interview, children were asked to discuss the questionnaires they had filled out (ALMERE, Godspeed, Empathy and NARS). Children felt that the questionnaires in general were easy to answer. They only identified that the original Perceived Safety dimension from the Godspeed questionnaire was not an easy question for them to answer since their emotional state changed during the interaction. So, this questionnaire was refined and now children are asked to rate their emotional state both at the beginning and at the end of the interaction. This final version of the questionnaire is presented in section 4.4.

6 References


Bryson, J. J. (2009). Robots Should Be Slaves. In Y. Wilks & J. Benjamins (Eds.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issue*


