EVALUATION REPORT

Pilot trainings for preschool teachers and Superhero Labs with children

ABOUT

This report evaluates the pilot phase of the Superhero Lab project: Engaging Children as Scientists to Explore Sustainability (2023-1-DE01-KA220-SN-23-36-151314). The evaluation covers two interconnected pilots conducted across six European partner countries, Austria, Germany, Greece, Lithuania, Slovenia, and Sweden, during 2025–2026: pilot trainings for preschool teachers, and Superhero Lab experiments with children aged 4–6.

June, 2026

Co-funded by the European Union

EVALUATION REPORT

Pilot trainings for preschool teachers and Superhero Labs with children

Project

The Superhero Lab: Engaging Children as Scientists to Explore Sustainability

Project number

2023-1-DE01-KA220-SN-23-36-151314

Prepared by

K&R Education Malmö AB on behalf of project consortium

On behalf of

Project consortium

Dissemination level

Public

Co-funded by the European Union

Contents

Executive summary

Purpose and context

This report evaluates the pilot phase of the Superhero Lab project: Engaging Children as Scientists to Explore Sustainability (2023-1-DE01-KA220-SN-23-36-151314). The evaluation covers two interconnected pilots conducted across six European partner countries, Austria, Germany, Greece, Lithuania, Slovenia, and Sweden, during 2025–2026: pilot trainings for preschool teachers, and Superhero Lab experiments with children aged 4–6.

Participants

A total of 153 preschool teachers participated in the pilot trainings, exceeding the project target of 120. Of these, 148 completed the post-training survey. A total of 767 children participated in the Superhero Lab activities, exceeding the target of 240. Every partner country met or exceeded both targets.

Teacher training: key findings

Satisfaction with the training was near-universal: 135 of 148 respondents (91.2%) gave the highest possible score of 5, and no respondent in any country rated satisfaction below 4. Reported gains in knowledge, confidence, and motivation were consistently high across all six countries. The most notable country difference was in self-assessed GreenComp familiarity, where Sweden scored substantially lower than other countries. As the evaluation argues, this reflects deeper engagement with the concept during the Swedish training rather than lower competence: teachers who spend more time on GreenComp tend to develop a more realistic awareness of the gap between their current knowledge and the full scope of the framework. These findings suggest that GreenComp functions primarily as a framework for curriculum developers and teacher educators, whereas preschool teachers engage with sustainability mainly through national curricula and practical teaching materials.

Superhero Labs with children: key findings

Children's self-assessments were consistently positive across all six countries and all six questions. The happy face was always the most frequently selected response. Enjoyment of working together showed no statistically significant country difference, confirming that the collaborative structure of the activities produces consistent outcomes regardless of local context. Children's experience of scientific learning showed greater variation: a 29.9 percentage point gap between the highest and lowest country on the nature learning question reflects differences in experiment choice and facilitation quality, not differences in children's capacity to learn. The Slovenian data illustrates this well: a higher rate of children reporting that the experiment did not turn out as expected reflects hypothesis-testing pedagogy, where children formed explicit predictions and reported honestly when predictions were wrong. In Austria, focus group feedback also indicated that the experiments supported language development in a linguistically diverse classroom, as children with limited German proficiency used the activities to communicate, describe observations, and practise new vocabulary.

Limitations

The evaluation relies on self-report measures, which may not fully capture actual competence or learning. Partner countries implemented different experiments, which may account for some of the cross-country variation in children's responses. Implementation quality was controlled partially, and the evaluation includes no pre-assessment baseline and no long-term follow-up data.

Areas for development

The evaluation identifies six areas for strengthening the programme: differentiating GreenComp content by audience; ensuring consistent minimum GreenComp coverage across partner countries; embedding more exercises on inquiry based pedagogy into the teacher training materials; providing more teacher training exercises on adapting activities for different group settings, such as large groups, mixed age groups, or groups with limited language proficiency; standardising the self-assessment instrument to black-and-white faces; and developing materials for children below the 4–6 year target range.

Actions taken

In response to the pilot findings, the teacher training programme was substantially revised. The revised two-day programme places participants in the learner role first, and includes hands-on experiment stations, hypothesis voting, video analysis, adaptation exercises, and a full experiment design task on Day 2. A structured roll-out model was also developed. Both the revised training with integrated trainer notes and the roll-out model are freely available at www.superherolab.eu.

Conclusion

The evaluation demonstrates that the Superhero Lab is both feasible and transferable across diverse European preschool contexts. While implementation quality depends on teacher facilitation and local conditions, the consistently high engagement of both teachers and children indicates strong potential for wider adoption. The revisions made following the pilot further strengthen the approach and provide a solid foundation for future dissemination and implementation.

1. Introduction

This evaluation report presents the results of the pilot phase of the Superhero Lab project, focusing on two interconnected activities: the training of preschool teachers and the implementation of Superhero Lab experiments with children. Both pilots were carried out across all six partner countries, Austria, Germany, Greece, Lithuania, Slovenia, and Sweden, during 2025–2026.

The Superhero Lab is a structured, experiment-based learning approach designed for preschool settings. It aims to engage children aged 4–6 years as active scientists, exploring themes from sustainable development through hands-on inquiry and collaboration. Preschool teachers are central to this approach: before leading activities with children, they participated in dedicated training sessions designed to build their competence, confidence, and motivation in science and sustainability education.

This report evaluates two aspects of the pilot:

- The pilot trainings for preschool teachers, evaluating whether the training strengthened their competence, confidence, and motivation to lead experiment-based learning.
- The Superhero Labs with children, evaluating children's engagement, learning, and experience during the experiments.

Data were collected through post-training surveys, individual written reflections, structured observations, and focus group interviews. The report presents the methods used, the results from each partner country including statistical analysis, and an integrated interpretation of the overall findings.

2. Purpose and objectives

Purpose

The purpose of this evaluation is to assess how the Superhero Lab material and approach function in authentic preschool settings, both as a tool for professional development of preschool teachers and as a learning method for children aged 4 to 6 years. The evaluation aims to generate evidence-based knowledge about the quality and effectiveness of both the teacher training and the children's activities, to inform the further development of the Superhero Lab materials and approach.

Objectives

Pilot trainings for preschool teachers

To conduct local pilot trainings to strengthen preschool teachers' competence, motivation, and confidence in leading experiment-based learning. Each partner was responsible for training at least 20 preschool teachers.

Superhero Labs with children

To carry out practical Superhero Labs with children aged 4–6 years, testing the learning materials in real educational contexts and collecting data on children's learning, engagement, and understanding of sustainability. Each partner was responsible for ensuring that at least 40 children participated in the activities.

Together, these two pilots contribute to the overall project by:

- Strengthening preschool teachers' competence in sustainability education and STEM-based learning.
- Providing children with opportunities to develop scientific thinking, critical reflection, and problem-solving skills.
- Ensuring that the learning materials are tested and evaluated in authentic learning environments.
- Generating a foundation for further development of the Superhero Lab method.

3. Method

3.1 Research approach and data collection

This evaluation is based on a combination of qualitative and quantitative methods to examine both pilots in preschool settings across the partner countries. The approach is descriptive and evaluative, aiming to capture both processes and outcomes from the perspectives of preschool teachers and children. All data collection was based on standardised project templates, ensuring comparability between countries.

The following data sources were used:

Post-training surveys

Preschool teachers completed structured self-assessments immediately after the training sessions. The surveys covered perceived understanding of GreenComp, motivation to conduct experiments, knowledge of leading science activities, and confidence in working with children. Responses were recorded on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree).

Individual written reflections

After implementing Superhero Lab activities with children, preschool teachers reflected individually on the implementation, children's responses, and their own professional development.

Structured observations of children

Preschool teachers documented children's participation using structured observation forms, focusing on engagement, communication, and collaboration.

Focus group interviews

Two focus group interviews were conducted in each partner country, the first after the initial training, and the second after teachers had implemented activities in practice, to capture professional change over time.

3.2 Implementation timeline

The pilot was implemented in all partner countries during 2025–2026 following a shared implementation plan. Trainings were led by facilitators who had previously participated in the project's Train-the-Trainer programme. Children's activities followed the project's structured experiment plans and included subsequent documentation and reflection.

3.3 Data analysis

The collected material was analysed in three steps:

Step 1 – Thematic analysis

Recurring patterns in qualitative data from teacher reflections, focus groups, and open-ended responses were identified and coded using thematic analysis (Braun & Clarke, 2012). Key themes include professional learning, motivation, confidence, and children's engagement.

Step 2 – Descriptive and inferential quantitative analysis

Likert survey responses were analysed by computing weighted means and standard deviations from the frequency distributions. Children's self-assessment responses were analysed as proportions of positive (happy-face) responses. Country differences were tested using chi-square tests of independence: for Likert data, scores were grouped into three categories (1–3, 4, and 5) to ensure adequate expected cell counts; for children's data, the original three response categories were used. Chi-square tests were selected because the survey responses were categorical and the analysis aimed to determine whether response distributions differed significantly between partner countries. Effect sizes are reported as Cramér's V (small: $V < 0.1$; medium: 0.1–0.3; large: $V > 0.3$).

Open-ended survey responses and focus group data were analysed using thematic analysis (Braun & Clarke, 2012), with themes identified inductively and then checked against the quantitative findings for consistency.

Step 3 – Integration

Quantitative and qualitative findings were interpreted together, with statistical results used to confirm, contextualise, or challenge the patterns identified in the qualitative analysis.

3.4 Limitations

Several limitations should be taken into account when interpreting the findings of this evaluation.

First, all outcome measures rely on self-report. Post-training surveys reflect participants' own perceptions of satisfaction, knowledge, and confidence rather than independently observed behaviour or competence. Children's self-assessments reflect felt experience rather than measured learning. Self-report instruments are subject to social desirability effects and may not fully capture what teachers learned or how children experienced the activities.

Second, partner countries implemented different Superhero Lab experiments. Because experiments vary in how directly they connect to nature and sustainability themes, some of the cross-country variation in children's responses, particularly on the nature learning question, may reflect experiment choice rather than differences in teaching quality or children's engagement.

Third, implementation quality varied in ways that were not systematically controlled. Differences in training depth, facilitation style, classroom size, and availability of support staff across countries mean that the results reflect the combined effect of the programme and its local context, not the programme in isolation.

Fourth, the evaluation does not include a pre-assessment. Survey data were collected only after the training, so it is not possible to quantify how much knowledge, confidence, or motivation changed as a result of participation. Reported scores reflect participants' state following the training but cannot be compared against a baseline.

Fifth, there was no long-term follow-up. Data were collected shortly after the training and the initial experiment sessions with children. Whether improvements in teacher confidence and motivation were sustained over time, remains unknown.

4. Participants

A total of **153 preschool teachers** participated in the pilot trainings across all six partner countries, **exceeding the project target of 120**. Of these, 148 completed the post-training survey; the remaining five did not submit a response.

Table 1

Number of preschool teachers per partner country

Country	Number of preschool teachers	Comment
Austria	22	As planned
Germany	23	As planned
Greece	27	As planned
Lithuania	36	As planned
Slovenia	20	As planned
Sweden	25	As planned

A total of **767 children participated** in the Superhero Lab activities, **exceeding the project target of 240**. In Sweden and Austria, younger children aged 1–3 years also participated as part of an extended local pilot beyond the primary target group of 4–6 year-olds.

Table 2

Number of children per partner country

Country	Number of children	Age	Comment
Austria	81	4–6 years	As planned. Included also 1–3-year-old children in one group.
Germany	127	4–6 years	As planned
Greece	138	4–6 years	As planned
Lithuania	119	4–6 years	As planned
Slovenia	77	4–6 years	As planned
Sweden	225	1–6 years	As planned. Included also 1–3-year-old children in some groups.

5. Results: Pilot trainings for preschool teachers

This section presents results from the evaluation of the pilot trainings, drawing on post-training surveys (n = 148), written reflections, and focus group interviews. Survey results are reported both as frequency distributions (Tables 3–9) and as means and standard deviations (Table 10), followed by statistical tests of country differences (Table 11).

5.1 Overall satisfaction and relevance

Satisfaction with the training as a whole was the highest-scoring and most uniform question in the entire survey. Overall, **135 of 148 respondents (91.2%) gave the highest possible score of 5**, with the remaining 13 (8.8%) giving a score of 4. **No respondent rated their satisfaction below 4 in any country**. The overall mean was 4.91 (SD = 0.28), the lowest standard deviation of any question, reflecting near-universal agreement that the training met expectations. Slovenia achieved a perfect mean of 5.00 (SD = 0.00), with all 20 respondents selecting the top score.

In the open-ended responses, the practical elements of the training were particularly valued. Participants across countries emphasised that concrete hands-on demonstrations made it easier to understand how to implement the experiments and transfer them into their own preschool practice.

5.2 Understanding of the GreenComp framework

Understanding of GreenComp, the European sustainability competence framework underpinning the Superhero Lab, shows the most striking country difference in the dataset. Overall, the mean was 4.28 (SD = 1.07), but interpreting this figure requires attention both to what the survey question actually measures and to what role GreenComp was intended to play in the training.

GreenComp was included in the training to build awareness of the sustainability competence framework that informed the design of the Superhero Lab experiments, **not as content that teachers were expected to master**. This was an intentional design choice. Preschool teachers work from national curricula, not EU policy documents, and GreenComp becomes relevant to their practice primarily when it is reflected in the frameworks and materials they use. From this perspective, the training's relationship to GreenComp was appropriate: it raised awareness rather than aiming for mastery. It is also directly relevant for teacher trainers and for the developers of learning materials, including the Superhero Lab experiments themselves, which were deliberately designed in alignment with GreenComp. These findings suggest that **GreenComp functions primarily as a framework for curriculum developers and teacher educators, whereas preschool teachers engage with sustainability mainly through national curricula and practical teaching materials**.

Against this background, Sweden's low scores (M = 2.88, SD = 1.11) may paradoxically reflect a more thorough engagement with the concept during training rather than less competence. Swedish trainers devoted more time to GreenComp, which raised more questions, more critical reflection, and a more accurate self-assessment of the gap between participants' existing knowledge and the full scope of the framework. In the other five countries, where GreenComp was mentioned more briefly, respondents may have given themselves higher ratings precisely because they had less insight into what they did not yet know. In Greece, Austria, Germany, and Lithuania, where between 87% and 100% of respondents gave top-two scores, the higher ratings may reflect lighter treatment of the concept rather than deeper understanding, a well-documented effect in self-assessment research where limited exposure to a topic makes it easier to feel confident about it.

A chi-square test confirms that the country differences are highly significant ($\chi^2(10) = 84.35$, $p < .001$), the largest statistical effect in the teacher survey. But the most important finding may not be the one that appears most obvious in the data. Rather than concluding that Sweden underperformed on GreenComp, the more productive reading is that the training's handling of GreenComp was inconsistent in depth and focus across partner countries, and that the survey question, as worded, measured something closer to self-perceived familiarity after the training session than to actual competence or understanding. Standardising the depth of GreenComp content across all partner trainings, and distinguishing more explicitly between awareness (appropriate for teachers) and mastery (appropriate for trainers and material developers), would produce more comparable and more interpretable data in future iterations.

5.3 Motivation to conduct experiments

Motivation to carry out the experiments with children was the second-highest scoring question overall (M = 4.82, SD = 0.62), with 131 of 148 respondents (88.5%) giving the top score of 5. The country differences are statistically significant ($\chi^2(10) = 25.70$, $p = .004$), but the pattern is unusual: rather than reflecting a generally low-performing country, it is driven by a small number of outlying responses in Austria.

In Austria, 3 of 22 respondents (13.6%) gave the lowest possible score of 1 on this question, while the remaining 19 all scored 5. This bimodal pattern is the main source of the statistically significant country difference, producing a mean of 4.45 (SD = 1.37). The three low scores are best understood in context: these respondents were educators working with children aged 1–3 years, a group that joined the Austrian pilot beyond the primary target range of 4–6 year-olds. The standard Superhero Lab experiments are designed for older children, and teachers of younger groups understandably felt less confident that the activities could be adapted to their setting without further support. This is not a reflection on the training quality but on the need for dedicated adaptation guidance for educators working with children below the intended age range. In all other countries, and among the remaining Austrian respondents, motivation scores were high and uniform, with means between 4.75 and 4.95.

5.4 Knowledge, confidence, and self-confidence

Three questions address dimensions of professional competence directly: knowledge of conducting experiments, confidence in carrying them out, and self-confidence in leading them. All three show high overall means (4.60–4.69) and no statistically significant country differences for knowledge ($\chi^2(10) = 13.89$, $p = .178$) or confidence ($\chi^2(10) = 10.32$, $p = .413$), suggesting that these outcomes were achieved consistently across the partnership regardless of context.

Self-confidence in leading experiments is the exception, showing a statistically significant country difference ($\chi^2(10) = 26.66$, $p = .003$). Austria achieved a perfect mean of 5.00 (SD = 0.00), all 22 respondents selected the top score, the only such result on any question in the survey. Sweden had the lowest mean (4.20, SD = 0.75), with 5 of 25 respondents (20.0%) giving a score of 3 and no respondent giving a 4 or 5 other than those directly mentioned. The gap between Austria and Sweden on this question (0.80 points) is consistent with the broader pattern of Swedish respondents reporting less familiarity with the GreenComp framework.

For all three dimensions, top-two scores (4 or 5) were given by between 88% and 100% of respondents in every country, indicating that the training was broadly effective at developing practical professional competencies regardless of starting point.

5.5 Survey results by country

Tables 3 to 9 present the full frequency distributions for all six survey questions, broken down by partner country. Table 10 summarises the means and standard deviations across countries, and Table 11 reports chi-square test results. In all frequency tables, the Likert scale runs from 1 (strongly disagree) to 5 (strongly agree).

Overall (all partner countries, n = 148)

Table 3

Post-training survey results, all partner countries (n = 148)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				13	135	148
My understanding of GreenComp has improved compared to before the training.	6	5	19	29	89	148
I feel motivated to carry out the experiments with the group of children.	3			14	131	148
My knowledge of conducting experiments has improved compared to before the training.			9	28	111	148
I feel more confident in carrying out experiments with the children than before the training.			8	38	102	148
My self-confidence in leading experiments with children has increased compared to before the training.			11	37	100	148

Greece (n = 22)

Table 4

Post-training survey results, Greece (n = 22)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				2	20	22
My understanding of GreenComp has improved compared to before the training.			2	6	14	22
I feel motivated to carry out the experiments with the group of children.				1	21	22
My knowledge of conducting experiments has improved compared to before the training.			1	1	20	22
I feel more confident in carrying out experiments with the children than before the training.			1	6	15	22

Question	1	2	3	4	5	n
My self-confidence in leading experiments with children has increased compared to before the training.			1	8	13	22

Austria (n = 22)

Table 5

Post-training survey results, Austria (n = 22)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				3	19	22
My understanding of GreenComp has improved compared to before the training.				3	19	22
I feel motivated to carry out the experiments with the group of children.	3				19	22
My knowledge of conducting experiments has improved compared to before the training.				3	19	22
I feel more confident in carrying out experiments with the children than before the training.				3	19	22
My self-confidence in leading experiments with children has increased compared to before the training.					22	22

Germany (n = 23)

Table 6

Post-training survey results, Germany (n = 23)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				2	21	23
My understanding of GreenComp has improved compared to before the training.	3				20	23
I feel motivated to carry out the experiments with the group of children.				2	21	23
My knowledge of conducting experiments has improved compared to before the training.			2	2	19	23
I feel more confident in carrying out experiments with the children than before the training.			2	3	18	23
My self-confidence in leading experiments with children has increased compared to before the training.			2	3	18	23

Lithuania (n = 36)

Table 7

Post-training survey results, Lithuania (n = 36)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				1	35	36
My understanding of GreenComp has improved compared to before the training.			3	9	24	36
I feel motivated to carry out the experiments with the group of children.				4	32	36
My knowledge of conducting experiments has improved compared to before the training.			2	8	26	36
I feel more confident in carrying out experiments with the children than before the training.			1	12	23	36
My self-confidence in leading experiments with children has increased compared to before the training.			3	11	22	36

Slovenia (n = 20)

Table 8

Post-training survey results, Slovenia (n = 20)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.					20	20
My understanding of GreenComp has improved compared to before the training.			2	9	9	20
I feel motivated to carry out the experiments with the group of children.				5	15	20
My knowledge of conducting experiments has improved compared to before the training.			1	7	12	20
I feel more confident in carrying out experiments with the children than before the training.			1	7	12	20
My self-confidence in leading experiments with children has increased compared to before the training.				5	15	20

Sweden (n = 25)

Table 9

Post-training survey results, Sweden (n = 25)

Question	1	2	3	4	5	n
I am satisfied with the training as a whole.				5	20	25
My understanding of GreenComp has improved compared to before the training.	3	5	12	2	3	25

Question	1	2	3	4	5	n
I feel motivated to carry out the experiments with the group of children.				2	23	25
My knowledge of conducting experiments has improved compared to before the training.			3	7	15	25
I feel more confident in carrying out experiments with the children than before the training.			3	7	15	25
My self-confidence in leading experiments with children has increased compared to before the training.			5	10	10	25

Summary statistics and country comparisons

Table 10

Means and standard deviations for post-training survey questions, by country

Question	Greece	Austria	Germany	Lithuania	Slovenia	Sweden	Overall
Satisfied with training	4.91 (0.29)	4.86 (0.34)	4.91 (0.28)	4.97 (0.16)	5.00 (0.00)	4.80 (0.40)	4.91 (0.28)
GreenComp understanding improved	4.55 (0.66)	4.86 (0.34)	4.48 (1.35)	4.58 (0.64)	4.35 (0.65)	2.88 (1.11)	4.28 (1.07)
Feel motivated to conduct experiments	4.95 (0.21)	4.45 (1.37)	4.91 (0.28)	4.89 (0.31)	4.75 (0.43)	4.92 (0.27)	4.82 (0.62)
Knowledge of conducting experiments improved	4.86 (0.46)	4.86 (0.34)	4.74 (0.61)	4.67 (0.58)	4.55 (0.59)	4.48 (0.70)	4.69 (0.58)
Feel more confident carrying out experiments	4.64 (0.57)	4.86 (0.34)	4.70 (0.62)	4.61 (0.54)	4.55 (0.59)	4.48 (0.70)	4.64 (0.58)
Self-confidence in leading experiments	4.55 (0.58)	5.00 (0.00)	4.70 (0.62)	4.53 (0.64)	4.75 (0.43)	4.20 (0.75)	4.60 (0.62)

Note. Values shown as M (SD). Scale: 1 = strongly disagree, 5 = strongly agree.

Table 11

Chi-square test results for post-training survey questions (country comparisons)

Question	χ^2	df	p	Significance
Satisfied with training	8.12	5	.150	ns
GreenComp understanding improved	84.35	10	< .001	***
Feel motivated to conduct experiments	25.70	10	.004	**
Knowledge of conducting experiments	13.89	10	.178	ns
Feel more confident carrying out exps	10.32	10	.413	ns
Self-confidence in leading experiments	26.66	10	.003	**

Note. Scores grouped into three categories (1–3, 4, 5) prior to testing. ns = not significant; ** p < .01; *** p < .001.

5.6 Focus group findings: teacher training

The focus group interviews conducted shortly after the training sessions or Superhero Labs with children reinforced and deepened the survey findings. In all six countries, teachers described the training as practically oriented and well-structured. The step-by-step experiment instructions and visual support materials were consistently mentioned as the most useful features, particularly for teachers who had limited prior experience with science activities.

The pattern of GreenComp scores in the survey was mirrored in the focus group data. Swedish teachers described GreenComp as a concept they had encountered and discussed at length during training, but found it abstract in relation to their day-to-day practice. Several Swedish participants noted that they were comfortable with sustainability as a general theme in their work, but that the specific competence categories of GreenComp did not map easily onto the national curriculum framework they actually use. This is consistent with the broader point that **GreenComp awareness among teachers is only as actionable as the degree to which their national curricula reflect it**. In other countries, where GreenComp was covered more briefly, few participants mentioned it specifically in the focus groups, which is itself telling: a concept introduced lightly tends to leave a lighter trace.

The focus groups also surfaced important contextual differences between partner countries that are not visible in the quantitative data. In Greece, preschool teachers frequently work with groups of more than 20 children without a classroom assistant, a structural condition that is notably more demanding than in most other partner countries, where teaching assistants are routinely present. This directly affects how much a teacher can do beyond managing the group: eliciting children's hypotheses, facilitating post-experiment discussion, and explicitly connecting the activity to learning concepts all require attention and time that is much harder to allocate when one adult is responsible for a large group alone. These structural differences should be taken into account when comparing outcomes across countries and when designing future implementation support.

6. Results: Superhero Labs with children

This section presents results from the implementation of the Superhero Lab experiments, drawing on structured observations, children's self-assessments (Tables 12–18), positive response rate comparisons (Table 19), chi-square tests (Table 20), teacher open-ended reflections (Tables 21–24), and focus group interviews.

6.1 Experiments conducted

All six partner countries implemented Superhero Lab experiments during the pilot phase. In all countries, the minimum target of 40 participating children was exceeded. Two experiments listed in the project materials, Pirate's Fatal Mistake (Oil Spill) and From Rain to Flood, were not conducted during this pilot.

Table 12

Superhero Lab experiments conducted during the pilot, by country

Nr	Experiment	Austria	Germany	Greece	Sweden	Slovenia	Lithuania
1	Solving Dried Markers	22	23	27	25	39	
2	Diaper Absorbents			20		33	
3	Three Little Pigs Are Building a Sustainable House	22					
5	Lemon Bubbly Experiment	22	23	27	50	33	36
6	Solar Pizza Oven			20			

Nr	Experiment	Austria	Germany	Greece	Sweden	Slovenia	Lithuania
7	Paper Recycling			20			
8	The Adventure of Droplet Watty			20	50		
10	Sleeping Daisy			20	10		
11	Give New Colour to Your Clothes	22	23	20			
12	Fabric Detectives			20			
13	Paper Rocket	22	46	27	25	39	36
14	Discovering Saw Bugs and Their Homes			20	10		
15	Secret Drawings with Invisible Ink	22		27	30	39	36

6.2 Children's self-assessments

After each experiment session, children rated their experience using a three-option self-assessment (happy face, neutral face, sad face). Tables 12 to 18 present the raw frequency counts per question and per country. Table 19 recasts these as positive response rates (percentage of children selecting the happy face), enabling clearer cross-country comparison. Table 20 reports chi-square tests for each question.

Overall (all partner countries)

Table 13

Children's self-assessment results, all partner countries

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	704	52	17
I thought it was fun to do the experiments together with the others.	622	37	14
We helped each other in the group when we did the experiment.	676	76	21
Did the experiment turn out the way you expected?	637	87	49
Did you learn anything about nature during the experiment?	610	80	77
I want to do more experiments.	726	27	14

Greece

Table 14

Children's self-assessment results, Greece

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	114	17	7

Question / Statement	Happy face	Neutral face	Sad face
I thought it was fun to do the experiments together with the others.	124	7	7
We helped each other in the group when we did the experiment.	116	17	5
Did the experiment turn out the way you expected?	114	17	7
Did you learn anything about nature during the experiment?	114	17	7
I want to do more experiments.	132	3	3

Austria

Table 15

Children's self-assessment results, Austria

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	82	4	1
I thought it was fun to do the experiments together with the others.	79	6	2
We helped each other in the group when we did the experiment.	80	4	3
Did the experiment turn out the way you expected?	72	11	4
Did you learn anything about nature during the experiment?	77	2	2
I want to do more experiments.	68	10	3

Germany

Table 16

Children's self-assessment results, Germany

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	113	11	3
I thought it was fun to do the experiments together with the others.	117	10	
We helped each other in the group when we did the experiment.	94	30	3
Did the experiment turn out the way you expected?	108	18	1
Did you learn anything about nature during the experiment?	101	5	21
I want to do more experiments.	117	7	3

Lithuania

Table 17

Children's self-assessment results, Lithuania

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	113	4	2
I thought it was fun to do the experiments together with the others.	113	5	1
We helped each other in the group when we did the experiment.	99	17	3
Did the experiment turn out the way you expected?	105	10	4
Did you learn anything about nature during the experiment?	115	3	1
I want to do more experiments.	116	2	1

Slovenia

Table 18

Children's self-assessment results, Slovenia

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	72	3	2
I thought it was fun to do the experiments together with the others.	70	6	1
We helped each other in the group when we did the experiment.	67	5	5
Did the experiment turn out the way you expected?	48	11	18
Did you learn anything about nature during the experiment?	53	18	6
I want to do more experiments.	68	5	4

Sweden

Table 19

Children's self-assessment results, Sweden

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?	210	13	2
I thought it was fun to do the experiments together with the others.	119	3	3
We helped each other in the group when we did the experiment.	220	3	2
Did the experiment turn out the way you expected?	190	20	15

Question / Statement	Happy face	Neutral face	Sad face
Did you learn anything about nature during the experiment?	150	35	40
I want to do more experiments.	225		

Positive response rates and country comparisons

Table 20

Percentage of positive (happy-face) responses to children's self-assessment questions, by country

Question	Greece	Austria	Germany	Lithuania	Slovenia	Sweden	Overall
How did it feel to do the experiment?	82.6%	94.3%	89.0%	95.0%	93.5%	93.3%	91.1%
Fun to do together with others	89.9%	90.8%	92.1%	95.0%	90.9%	95.2%	92.4%
We helped each other	84.1%	92.0%	74.0%	83.2%	87.0%	97.8%	87.5%
Experiment turned out as expected?	82.6%	82.8%	85.0%	88.2%	62.3%	84.4%	82.4%
Did you learn anything about nature?	82.6%	95.1%	79.5%	96.6%	68.8%	66.7%	79.5%
I want to do more experiments	95.7%	84.0%	92.1%	97.5%	88.3%	100%	94.7%

Table 21

Chi-square test results for children's self-assessment questions (country comparisons)

Question	χ^2	df	p	Cramér's V	Effect size
How did it feel to do the experiment?	20.14	10	.028	0.114	Medium
Fun to do together with others	15.04	10	.131	ns	—
We helped each other	61.13	10	< .001	0.199	Medium
Experiment turned out as expected?	52.24	10	< .001	0.184	Medium
Did you learn anything about nature?	85.69	10	< .001	0.236	Medium
I want to do more experiments	44.64	10	< .001	0.171	Medium

Note. Tests used the three original response categories (happy, neutral, sad). ns = not significant; some cells had expected frequencies below 5 and results should be interpreted with caution.

Interpretation of children's self-assessment results

The overall pattern is strongly positive: across all six questions and all six countries, the happy-face response was always the most frequently selected. However, the statistical tests reveal that country differences are significant and meaningful for five of the six questions, with all effect sizes falling in the medium range (V = 0.114–0.236). The one question that shows no significant country difference is

enjoyment of working together with others ($\chi^2(10) = 15.04$, $p = .131$), indicating that the collaborative dimension of the activities is consistently experienced across all settings, regardless of context.

How it felt to do the experiment shows the smallest significant effect ($\chi^2(10) = 20.14$, $p = .028$, $V = 0.114$). Greece recorded the lowest positive rate at 82.6% (114 of 138 children), compared to Lithuania at 95.0% (113 of 119). The difference is real but moderate, and may partly reflect the structural conditions of Greek preschools, where one teacher works with more than 20 children without an assistant, leaving less capacity to attend to each child's individual experience of the activity.

Collaboration ("We helped each other") shows a clear and large split ($\chi^2(10) = 61.13$, $p < .001$, $V = 0.199$). Germany recorded the lowest positive rate at 74.0% (94 of 127 children), with 30 children (23.6%) giving a neutral response, more than three times the neutral rate in Sweden (3 of 225, 1.3%). Sweden achieved the highest collaboration rate in the entire dataset at 97.8% (220 of 225 children). This contrast is one of the sharpest country differences in the data and likely reflects differences in how the experiments were structured and facilitated rather than differences in children's willingness to work together.

Whether the experiment turned out as expected requires careful interpretation ($\chi^2(10) = 52.24$, $p < .001$, $V = 0.184$). In Slovenia, 18 of 77 children (23.4%) gave a sad-face response, the highest rate of any country on any question in the entire self-assessment dataset. At first glance this appears to be the most negative result in the data. In reality it is the opposite. The experiment in question in Slovenia was Solving Dried Markers, and the Slovenian teacher conducted the activity with voting: children's hypotheses were collected and voted on before the experiment began, making the outcome explicitly testable against a prior prediction. When the result differed from what children had expected, they reported this honestly. This is precisely what scientific inquiry requires, and it means that Slovenia's 23.4% sad-face rate is not a sign that the experiment failed, but a sign that children were thinking scientifically. In countries where no explicit hypothesis was elicited, children had no expectation to compare against, making it easier to select the happy face regardless of what happened. Lithuania's sad-face rate on the same question was just 3.4% (4 of 119 children), but this likely reflects how the activity was facilitated.

Learning about nature shows the largest effect size in the children's data ($\chi^2(10) = 85.69$, $p < .001$, $V = 0.236$). Positive rates range from 66.7% in Sweden (150 of 225 children) and 68.8% in Slovenia (53 of 77), to 96.6% in Lithuania (115 of 119) and 95.1% in Austria (77 of 81). In Sweden, 40 children (17.8%) gave a sad-face response, more than four times the number in Lithuania (1 child, 0.8%). This variation is not straightforwardly a difference in how much children learned. Two concrete factors are likely to explain much of it. First, experiment choice: some activities have a more immediately visible connection to nature (such as The Adventure of Droplet Watty or Sleeping Daisy) while others, like Paper Rocket or Solving Dried Markers, require more deliberate pedagogical bridging to make the nature learning explicit. Countries that selected experiments with a less obvious nature connection would be expected to produce lower scores on this question. Second, how much time teachers devoted to structured observation, discussion, and reflection after each experiment: when teachers explicitly helped children articulate what they had learned about the natural world, children were more likely to experience the activity as learning about nature. When the experiment was experienced primarily as a hands-on activity without a structured debrief, the scientific content was less salient.

Wanting to do more experiments was the most positive question overall, with a mean positive rate of 94.7% and Sweden achieving a 100% rate (all 225 children). Austria recorded the lowest rate at 84.0% (68 of 81 children, $\chi^2(10) = 44.64$, $p < .001$, $V = 0.171$). However, the Austrian data on this question must be read alongside an important methodological finding: during the pilot, it was observed that some Austrian children appeared to select a response based on the color of the smiley face rather than their actual experience, choosing the red face because red was their favorite color, for instance. Austria responded by switching to black-and-white printed smiley faces during the pilot, which means part of the Austrian data was collected with the colored instrument and part with the revised version. It is not known whether other partner countries were affected by the same issue, as no systematic check was conducted. This finding has direct implications for the validity of the instrument as currently designed. For children aged 4–6, color is a particularly salient perceptual attribute, and if color preference influences response selection, the data may not reliably reflect children's actual experience. Standardising to black-and-white faces across all countries

in future iterations is a straightforward methodological improvement that would strengthen the comparability and reliability of the tool.

6.3 Children's open-ended responses

The open-ended responses from children show recurring descriptions of positive experiences across all countries. The most common themes are that the activities were fun, exciting, and something children wanted to repeat. Children frequently describe conducting the experiments together, helping each other, sharing materials, and working through steps as a group. In several countries, children also asked follow-up questions about the phenomena they observed, suggesting the experiments sparked broader curiosity beyond the session itself.

6.4 Teacher reflections on the children's activities

After implementing the experiments, 49 preschool teachers completed individual written reflections covering four areas: design and materials, practical use, learning and exploration, and personal professional development. The results are presented in Tables 22 to 25.

Design and materials (N = 49)

Table 22

Design and materials: teacher reflection responses (N = 49)

Statement	Number of ticks
The material was clear and easy for me to understand	49
The experiment encouraged active participation from the children and sparked their curiosity	45

Practical use (N = 49)

Table 23

Practical use: teacher reflection responses (N = 49)

Statement	Number of ticks
The children showed interest in the experiment	46
The activity supported the children's critical thinking, exploration, and STEM learning	23

Learning and exploration (N = 49)

Table 24

Learning and exploration: teacher reflection responses (N = 49)

Statement	Number of ticks
The children demonstrated curiosity and asked questions during the experiment	49
The activity supported children's critical thinking, exploration, and learning in STEM (Science, Technology, Engineering, Mathematics)	42

Personal reflection (N = 49)

Table 25

Personal reflection: teacher reflection responses (N = 49)

Statement	Number of ticks
I have learned more about leading STEM-related experiments	36
I feel more confident working with sustainable learning in the children's group	39

All 49 teachers confirmed that the material was clear and easy to understand, and all 49 confirmed that children demonstrated curiosity and asked questions during the experiment, the two questions that achieved unanimous agreement. The statement about STEM learning support shows the largest gap: 42 of 49 teachers (85.7%) affirmed this in the Learning and Exploration section, while only 23 of 49 (46.9%) confirmed it in the Practical Use section. The difference in wording between these two statements is important: the Practical Use question is more concrete ("critical thinking, exploration, and STEM learning"), while the Learning and Exploration question adds "curiosity and asked questions" as a separate element, which may have made it easier to affirm. Taken together, the responses suggest broad but not universal perception that the activities supported children's scientific thinking.

On personal professional development, 36 of 49 teachers (73.5%) indicated they had learned more about leading STEM-related experiments, and 39 of 49 (79.6%) felt more confident in sustainable learning. The gap between these figures and the 100% agreement on children's engagement may indicate that some teachers already brought sufficient prior competence that the activities did not produce new personal learning, even as they clearly engaged children effectively.

6.5 Focus group findings: children's activities

Focus group interviews after the implementation were consistent across countries in their overall picture: teachers in all six countries described children as curious, active, and engaged throughout the experiments, asking questions, formulating hypotheses, testing ideas, and reflecting on results together. This qualitative picture aligns closely with the quantitative self-assessment data, where the overall happy-face rate across all questions was above 79%.

The pattern of country variation identified statistically was also visible in the focus groups. German teachers described some groups finding the collaboration aspect challenging at first, which is consistent with Germany's lower collaboration positive rate (74.0%). Slovenian teachers described children's genuine surprise at the outcome of the Solving Dried Markers experiment, where results differed from what children had predicted, a finding that, as discussed in Section 6.2, reflects hypothesis-testing pedagogy rather than a problem with the activity. Swedish and Slovenian teachers both raised the difficulty of explicitly connecting the experiments to learning about nature, which corresponds to these countries' lower positive rates on that question (66.7% and 68.8% respectively).

The focus groups surfaced important contextual differences between partner countries. In Greece, preschool teachers typically work with groups of more than 20 children without a classroom assistant, a structural condition that is materially more demanding than in most other partner countries, where teaching assistants are routinely present. This directly limits how much attention a teacher can give to deepening learning, eliciting hypotheses, and facilitating post-experiment reflection when the teacher's primary responsibility is also managing the group. These structural differences should be taken into account when comparing outcomes across countries, and should inform how future implementation support is designed.

The focus groups also revealed differences in how deeply different teachers engaged with the learning dimension of the activities. In some cases, experiments were experienced primarily as enjoyable hands-on activities, with less emphasis on observation, hypothesis formation, or post-experiment discussion. This is the evaluators' professional interpretation of the patterns observed rather than a direct finding from the

focus group transcripts; it should be treated as a hypothesis for further investigation. What the focus group data does show clearly is that the sessions generating the most learning-focused discussion were those where teachers had prepared specific questions in advance and explicitly planned time for structured reflection after the activity.

An Austrian teacher described implementing the experiments in a group where several children had limited proficiency in German. The teacher observed that the activities created natural opportunities for communication, as children discussed observations, asked questions, and explained what they were doing. According to the teacher, repeated participation in the experiments appeared to support the children's development of German language skills alongside their scientific learning.

The materials were widely described as flexible and adaptable, one of the most consistently cited strengths across all focus groups. At the same time, preparation time, material availability, and the need for clearer documentation guidance were recurring concerns. Teachers in several countries also requested more explicit pedagogical support for facilitating the scientific learning dimension of the experiments, guidance on how to frame what children are discovering, how to respond to unexpected outcomes, and how to help children connect the experiment to concepts about nature and sustainability. This was later incorporated into the teacher training materials.

7. Analysis

The analysis draws on thematic analysis (Braun & Clarke, 2012) integrated with the statistical findings from the survey and self-assessment data. Four themes emerged consistently across all data sources.

Training outcomes are strong but require differentiated interpretation

The teacher training achieved its primary objectives on most dimensions. For satisfaction, knowledge, confidence, and motivation, the vast majority of respondents in every country gave top-two scores (88–100% per country), and statistical tests reveal no significant country differences on knowledge ($\chi^2(10) = 13.89$, $p = .178$) or general confidence ($\chi^2(10) = 10.32$, $p = .413$). These dimensions can be considered robustly and consistently achieved across the partnership.

The GreenComp finding requires more layered interpretation than the raw numbers suggest. Taken at face value, Sweden's mean of 2.88 appears to indicate that the training failed to develop GreenComp understanding in that country. In reality, Sweden devoted more training time to GreenComp, which led to more questions, more honest self-assessment, and a more accurate recognition of the gap between current knowledge and the full scope of the framework. The countries with near-perfect GreenComp scores may have given the topic less time, leaving participants with less insight into what they did not yet know, a well-established dynamic in self-assessment research. The chi-square result ($\chi^2(10) = 84.35$, $p < .001$) confirms a real country difference, but the difference is in depth of engagement, not in outcome quality.

As discussed in Section 5.2, the expectation that teachers should score highly on GreenComp familiarity reflects a misalignment between the training's design intent and the survey question. GreenComp functions primarily as a framework for curriculum developers and teacher educators; preschool teachers engage with sustainability mainly through national curricula and practical materials. Measuring teachers' GreenComp familiarity therefore captures a distal rather than proximal outcome of the training. A more relevant indicator for future evaluations would be whether teachers can connect the experiments they conduct to sustainability learning goals in their national curriculum.

Children's engagement is consistently high but uneven

Children's responses show that the Superhero Lab reliably creates positive, motivating learning experiences: **the happy face was always the modal response across all six questions and all six countries**, and the two dimensions most closely related to experience, how it felt, and wanting more experiments, both exceeded 90% positive overall. The social dimension showed no significant country difference ($\chi^2(10) = 15.04$, $p = .131$), indicating that the collaborative structure of the activities functions with notable

consistency across diverse settings. This robustness in the social and motivational dimensions is itself a meaningful finding: it suggests that **the Superhero Lab format can be trusted to create a positive group learning experience even when other dimensions are more variable**.

The scientific learning dimension is more variable, but the nature of that variability matters. The "learning about nature" question shows both the largest effect size (V = 0.236) and the widest country range (66.7% to 96.6%), **a 29.9 percentage point gap**. This is not random variation. It is shaped by concrete factors: which experiments were chosen, and how much structure teachers provided for observation, discussion, and reflection. Countries where teachers explicitly connected the activity to what children were learning about the natural world produced higher scores; countries where the experiment was experienced primarily as a hands-on activity without structured debrief produced lower scores. This is not a finding about children's capacity to learn, it is a finding about what pedagogical conditions are needed to make that learning visible and felt, and about the degree to which the current materials provide sufficient guidance for achieving those conditions.

Slovenia's 23.4% sad-face rate, the highest rate of any country on any question, initially appears to be the most negative result in the data. It is in fact meaningful. The Slovenian teacher collected and voted on hypotheses before the Solving Dried Markers experiment, so children had a specific expectation to compare against when the result came in. When it differed from their prediction, they reported this honestly. This is the Superhero Lab functioning as intended: as an inquiry-based science experience in which children form hypotheses, test them, and engage with the outcome, including when the outcome surprises them. The contrast with Lithuania's 3.4% sad-face rate on the same question does not mean Lithuanian children had a better experience, it more likely means they were not asked to form an expectation beforehand, and so had nothing to be wrong about. The self-assessment instrument, as currently designed, is sensitive to this difference in facilitation depth.

Social and collaborative learning is the most robust outcome across diverse educational contexts

The one question showing no significant country difference is enjoyment of collaboration, and wanting to do more experiments, the most positive question overall, also shows only a medium effect size (V = 0.171) despite statistical significance. Together, these findings suggest that the Superhero Lab format reliably produces motivation to continue and positive experiences of working together, even where the scientific content or experimental outcomes are more variable.

This robustness is consistent with the sociocultural theoretical framework informing the approach (Vygotsky, 1978; Lave & Wenger, 1991): the structured collaborative format creates consistent social learning conditions regardless of local context, while the cognitive and content related dimensions are more sensitive to implementation quality, teacher preparation, and contextual factors. The collaborative nature of the activities not only supports children's social participation but also creates authentic opportunities for language development through observation, discussion, questioning, and shared problem solving. This was illustrated in the Austrian pilot, where a teacher reported that children with limited German proficiency developed their language skills through active participation in the experiments. The social dimension of the Superhero Lab can therefore be considered a reliable core outcome, while the scientific learning dimension remains an area of potential that depends more on teacher facilitation than the materials alone can ensure.

Implementation variation is real, statistically confirmed, and contextually grounded

Across both data sets, five of six children's questions and three of six teacher survey questions show significant country differences, all with medium effect sizes that are practically meaningful. This variation reflects real differences in how the same materials are used under different conditions: pedagogical culture, prior professional experience, classroom structure, experiment selection, teacher identity, and the depth of training engagement all play a role.

Structurally, classroom organisation differs substantially between partner countries, and this is directly relevant to outcomes. Greek preschool teachers typically manage groups of more than 20 children without

an assistant, a condition under which inquiry-based facilitation, hypothesis elicitation, and individual attention are materially harder to achieve than in settings where assistants are present. This likely contributes to Greece's lower positive rate on how it felt to do the experiment (82.6%) and may influence other dimensions where individual attention matters. These structural realities are outside the control of the Superhero Lab programme, but they should be acknowledged in how outcomes are interpreted and in how implementation support is designed for different national contexts.

The countries where children show lower rates of perceived nature learning (Sweden and Slovenia) are also the countries where teachers report either lower GreenComp familiarity or describe specific difficulty in making the scientific framing of the experiments explicit. This alignment between teacher and child data is not coincidental: when teachers are less certain about the sustainability and science framework behind the activities, this is transmitted, not through failure of effort, but through the depth of the learning conversations they are able to have with children. Strengthening teachers' confidence in the scientific and sustainability framing of the activities, through clearer pedagogical guidance embedded in the materials themselves, is therefore likely to be the most direct lever for improving children's experience of the learning dimension.

8. Future recommendations

The results, read together with the contextual information gathered during the pilot, identify seven specific areas where the Superhero Lab materials, training approach, and supporting instruments could be strengthened:

- Differentiated GreenComp integration by audience: GreenComp serves different purposes for different groups. For classroom teachers, awareness of the framework and its connection to preschool practice is sufficient. For teacher trainers and learning material developers, a deeper understanding is needed, since GreenComp directly informs what the activities are designed to develop. For policymakers and curriculum developers, the critical task is translating GreenComp into national curricula that teachers are required to follow, because it is only when sustainability competencies are embedded in national frameworks that teachers can act on them systematically. Future training design should reflect these distinctions explicitly, and advocacy work should accompany the programme to encourage decision makers to integrate GreenComp-aligned competencies into national early childhood curricula.
- Pedagogical bridging guidance for scientific learning: The 29.9 percentage point country gap on the nature learning question reveals that children's experience of scientific learning depends significantly on how the teacher facilitates observation, discussion, and reflection around the experiment. Teacher training focusing on guiding questions for before, during, and after the experiment; language for normalising unexpected outcomes; and reflection cues that invite children to name what they have learned about nature, would reduce dependence on individual teacher initiative and strengthen the learning dimension of the activities across all contexts.
- Structural adaptation guidance: The teacher training currently does not account for the significant differences in classroom structure between partner countries. A practical supplement addressing how to implement the activities in large-group settings without a classroom assistant, including modifications to group size, timing, and facilitation format, would support teachers working under more demanding structural conditions and reduce outcome inequality between settings.
- Standardisation of the self-assessment instrument: The Austrian pilot identified that some children selected their response based on the color of the smiley face rather than their experience. Austria switched to black-and-white faces during the pilot, but this means part of the Austrian data is not fully comparable with the rest. Whether other countries were affected is unknown. For future use, the instrument should be standardised to black-and-white faces across all countries, and the administration guide should include a brief protocol check confirming that children understand the faces represent feelings rather than colors.

- Preparation time and material accessibility: Multiple focus groups identified extensive preparation time and difficulty obtaining specific materials as practical barriers. Simplified preparation guides and lists of commonly available substitute materials would strengthen day-to-day usability without changing the pedagogical content.
- Documentation support and age adaptation: Requests for clearer documentation templates for piloting emerged in several countries. More structured recording sheets for capturing children's responses would reduce teacher workload and improve data consistency. Where younger children (1–3 years) participated, as in Sweden and Austria, teachers noted the need for adapted materials. An optional supplement with guidance for this extended age range could support the broader use observed in practice.

9. Conclusion

The pilot phase demonstrates that both the teacher training and the Superhero Labs with children have been successfully implemented across all six partner countries, with results that are broadly strong on both quantitative and qualitative measures. Reading those results carefully, however, reveals a more textured picture than the headline numbers suggest, one that points to specific and actionable directions for the next phase.

For the teacher training, satisfaction was near-universal (M = 4.91, SD = 0.28; 100% top-two scores in every country), and professional competencies in knowledge, confidence, and motivation were consistently strengthened.

For the Superhero Labs, social and motivational outcomes were strong and consistent across all countries. The collaborative nature of the activities not only promoted engagement and cooperation but also showed potential to support language development in multilingual classrooms. In the Austrian pilot, for example, a teacher reported that children with limited German proficiency developed their language skills through participation in the experiments. The scientific learning dimension showed more variation, as discussed in Sections 6.2 and 7: the 29.9 percentage point gap on the nature learning question reflects differences in experiment choice and facilitation depth, not children's capacity to learn. The Slovenian finding, where 23.4% of children gave a sad face response on expected outcomes, is the most scientifically meaningful result in the dataset for that question, reflecting authentic hypothesis testing pedagogy. The measurement issues identified in Austria, including the color preference effect on the self assessment instrument, point to practical improvements needed before the next data collection round.

The evaluation demonstrates that the Superhero Lab is **both feasible and transferable across diverse European preschool contexts**. While implementation quality depends on teacher facilitation and local conditions, the consistently high engagement of both teachers and children indicates **strong potential for wider adoption**. The revisions made following the pilot further strengthen the approach and provide a solid foundation for future dissemination and implementation.

10. What happened next: actions taken in response to the pilot

The pilot evaluation generated concrete evidence that directly informed two follow-up outputs: a substantially revised teacher training programme and a structured roll-out model. Both are available free of charge at www.superherolab.eu.

10.1 Revised teacher training

One of the clearest signals from the pilot was that professional development for preschool science education is most effective when it moves beyond demonstrating experiments. Teachers who understood the pedagogical purpose behind each step, why a hypothesis is collected before the experiment, what makes a question open rather than closed, how to respond when a child's prediction turns out to be wrong,

were better equipped to facilitate genuine inquiry with children. The revised training addresses this directly.

The training is structured as a two-day programme that places participants in the role of learners before asking them to reflect as educators. Experiments are first experienced hands-on, mixing invisible ink, voting on hypotheses, launching paper rockets, testing absorbency, before the pedagogical analysis begins. This mirrors the approach teachers are expected to use with children: curiosity and discovery before explanation.

The programme is built around active participation throughout both days. Key interactive features include:

- Icebreaker activities that model the same inquiry spirit expected in the preschool classroom, such as a Bingo card asking participants to find colleagues who match various descriptions, which mirrors how open-ended questioning can make a group curious about each other.
- Hands-on experiment station where participants experience material as a child would, with one minute to touch, smell and observe materials before any instruction is given, the same protocol recommended for use with children.
- Hypothesis voting exercises, including a whole-group chart recording YES/NO predictions before the experiment, and small-group comparison of results against predictions afterwards. Participants practice the exact facilitation moves they will later use with children.
- Video analysis tasks with a structured observation framework, guiding participants to notice inquiry moments, vocabulary opportunities, sustainability connections, safety considerations, and opportunities for age differentiation.
- Quiz activities throughout both days that check understanding of the science background, the pedagogical reasoning behind each step, and how to handle challenging situations such as a child's hypothesis being wrong, a group losing interest, or an experiment producing unexpected results.
- Triad work, consistent collaborative reflection in groups of three, in which participants discuss how each experiment could support development in their own learning environment, what adaptations would be needed, and how they would rate their own confidence in using the activity.
- Adaptation exercises where participants explicitly discuss how to modify the same experiment for children aged 1-3, for groups with limited language skills, and for mixed-age settings, addressing the structural diversity identified as significant in the pilot evaluation.
- Story-writing in triads: participants create their own short story introduction for an experiment (a character, a problem, and a question ending with "Can you help?"), directly practising the narrative technique used to open every Superhero Lab activity.
- A full experiment design task on Day 2, in which participants work in groups to design a new Superhero Lab experiment from scratch using the six-step structure, before presenting it to the group against a checklist covering child involvement, curiosity, prediction, discovery, and sustainability connection.

The training is accompanied by integrated trainer notes that guide facilitators through each activity, including timing guidance, discussion questions, suggested responses to common questions, and adaptation suggestions. It is available as a complete presentation with trainer notes, free of charge at www.superherolab.eu, alongside instructional videos for each of the 15 experiments on the SuperHeroLabProject YouTube channel.

10.2 Roll-out model

Based on the pilot results, including the evidence of what worked consistently across countries and what varied with implementation quality, a structured roll-out model document was developed to support the broader dissemination and sustained use of the Superhero Lab approach. The roll-out model addresses how partner organisations and new adopters can implement the training and the experiments in a way that is faithful to the methodology while remaining adaptable to local conditions. It is available free of charge at www.superherolab.eu.

11. Practical implications

The findings from this evaluation have concrete implications for different groups with a stake in early childhood science and sustainability education.

For preschool teachers

The pilot demonstrates that experiment-based science learning is not only feasible in preschool settings but consistently engaging for children across very different national contexts. Teachers do not need to be science specialists to lead these activities effectively, what matters is creating the conditions for children to ask questions, make predictions, and discover outcomes, rather than explaining science content in advance or demonstrating experiments.

The most important practical insight for teachers is that the depth of children's scientific learning is shaped by the quality of the conversation around the experiment, asking open-ended questions, eliciting predictions before the activity begins, and explicitly connecting what happened to what children are learning about nature, rather than by the experiment itself. Teachers who planned their facilitation questions in advance and built in time for structured reflection after the activity consistently produced richer learning experiences than those who focused primarily on the experimental procedure.

For teacher educators

The evaluation points clearly to a gap in conventional professional development for preschool science: training that focuses on demonstrating experiments does not sufficiently prepare teachers to facilitate scientific inquiry. Participants in this pilot who understood why each step of the Superhero Lab structure exists, what it means for a child to form a hypothesis, why an unexpected result is more valuable than a confirmed one, how to respond to a wrong prediction without closing down curiosity, were measurably more confident and more effective in their implementation.

Future training programmes should allocate substantial time to practising inquiry facilitation as a skill in its own right: rewriting closed questions into open ones, responding to children's incorrect explanations in ways that extend thinking, creating productive hypothesis-testing situations, and framing unexpected outcomes as scientific data rather than failures. The triad-based reflection format used in the revised Superhero Lab training provides a practical model for how this kind of facilitation practice can be embedded in professional development without requiring additional content knowledge from participants.

For curriculum developers

The evaluation shows that children's experience of scientific learning is most positive when the connection between an experiment and a sustainability concept is made explicit, not left for children to infer. Curriculum guidelines that translate sustainability competencies into specific examples of preschool practice (what questions to ask, what phenomena to observe, how to connect an experiment to a real-world environmental challenge) would directly support the kind of teaching observed in the most effective implementation contexts in this pilot.

GreenComp provides a valuable reference framework for what sustainability competencies look like at a conceptual level. What is currently missing, and what this pilot has begun to address through the Superhero Lab materials, is a layer of practical translation: concrete activities, questions, and pedagogical approaches that make those competencies accessible to teachers working with children aged 4–6. Curriculum developers have a role in producing and legitimising this translation layer. This translation into national curricula and teacher education is essential to ensure that inquiry based science education becomes a routine part of early childhood practice rather than depending on individual teachers' confidence, competence, or motivation.

For policymakers

The evaluation confirms a structural finding that cannot be addressed through teacher training alone: **GreenComp awareness among teachers is only as actionable as the degree to which national curricula reflect it**. Teachers in this pilot worked from national frameworks, not EU policy documents, and their ability to connect sustainability themes to children's learning depended on whether their national context gave them a mandate and a vocabulary for doing so.

As discussed in Sections 5.2 and 7, GreenComp awareness among teachers is closely linked to how well these competencies are reflected in the national curricula and materials that teachers use in practice. Policymakers responsible for early childhood education frameworks have an opportunity to embed sustainability competencies in expected learning outcomes, pedagogical guidelines, and planning frameworks. The Superhero Lab pilot demonstrates that when teachers are given both the materials and professional support to engage children in experiment-based sustainability learning, children respond with high engagement and curiosity. Creating the conditions for this kind of practice to become part of standard preschool provision, rather than a project-based exception, is where policymakers can have the most lasting influence.

12. References

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological (pp. 57–71). American Psychological Association.

Dewey, J. (1938). Experience and education. Macmillan.

European Commission. Improving science teaching and learning. https://cordis.europa.eu/article/id/91406-improving-science-teaching-and-learning [Retrieved 2026-05-26].

GO-Lab. GO Lab support. https://support.golabz.eu/sv/node/203 [Retrieved 2026-05-26].

Lave, J. & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). Science education now: A renewed pedagogy for the future of Europe. European Commission.

Skolverket. Lärportalen. https://larportalen.skolverket.se/api/resource/P03WCPLAR177267 [Retrieved 2026-05-26].

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

13. Appendices

Appendix 1: Guidelines for preschool teachers

This document is a guide for trained facilitators conducting Superhero Lab training for preschool teachers. It covers the training structure, the activity plan with children, and the individual reflection process for teachers.

Part 1: Guide for trained facilitators

Below are the project objectives that should form the basis for planning the training with the participating preschools:

Goal 1

Each partner country shall train at least 20 preschool teachers from participating preschools in planning and conducting Superhero Lab experiments.

Goal 2

At least 40 children aged 4-6 shall be given the opportunity to participate in at least two Superhero Lab experiments led by trained preschool teachers.

Training structure

Schedule one or more sessions for training with the participating preschools. The recommended duration is at least 1.5 hours. A PowerPoint presentation is available for use during the training.

Step 1: Introduction to the project

Use the PowerPoint presentation as your starting point. Provide an overview of the Superhero Lab project and show participants where they can find the videos and instructions for the different experiments.

Step 2: Demonstrate and try out

Demonstrate and conduct 2-3 experiments for the preschool teachers. Allow them to try performing the experiments in work groups. Choose experiments that best suit the preschool's activities and prepare all necessary materials in advance.

Step 3: Plan the implementation

The preschool teachers individually or in smaller groups plan how to conduct Superhero Lab experiments with their own groups of children, using the pedagogical planning format the preschool normally uses.

Step 4: Individual self-assessment of the training

Preschool teachers complete an individual self-assessment rated on a scale from 1 (does not agree at all) to 5 (fully agree):

- I am satisfied with the training as a whole.
- My understanding of GreenComp has increased compared to before the training.
- I feel motivated to conduct the experiments with the children's group.
- My knowledge of how to lead experiments has increased compared to before the training.
- I feel more confident in leading experiments with the children compared to before the training.
- My self-confidence in leading experiments with the children has increased compared to before the training.

Step 5: Submission

The completed reflection forms (A1.2 and A1.3) should be submitted to the responsible trainer at each preschool, then to the partner organisation. Submission deadlines: 1 October 2025, 1 December 2025, 2 February 2026, 20 March 2026.

Part 2: Activity plan with the children

Step 1: Conduct the experiment

Carry out the chosen experiment with the children's group based on the pedagogical planning prepared during the training session.

Step 2: Reflection with the children

Immediately after completing the experiment, reflect together with the children using A1.1 (self-assessment questions) and A1.2 (data collection form).

Part 3: Individual reflection for preschool teachers

Immediately after conducting the experiment and the reflection with the children's group, reflect individually based on your observations and dialogue with the children. Use A1.3.

A1.1: Self-assessment questions for children

Ask the children one question at a time and have them point to the face (happy, neutral, or sad) that best represents how they feel.

- How did it feel to do the experiment?
- I thought it was fun to do the experiments together with the others.
- We helped each other in the group when we did the experiment.
- Did the experiment turn out the way you expected?
- Did you learn anything about nature during the experiment?
- I want to do more experiments.

A1.2: Data collection form: children's responses

Fill in how many children gave each answer for each question or statement.

Superhero Lab experiment conducted: _____

Number of children who participated: _____

Country: _____

Table A1.1

Children's self-assessment response form (A1.2)

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?			
I thought it was fun to do the experiments together with the others.			
We helped each other in the group when we did the experiment.			
Did the experiment turn out the way you expected?			
Did you learn anything about nature during the experiment?			
I want to do more experiments.			

A1.3: Individual reflection form for preschool teachers

Reflect and respond individually. Tick the boxes if the statements apply and add comments in the free-text sections if you wish.

Experiment conducted: _____

Number of children who participated: _____

Country: _____

1. Design and materials

☐ The material was clear and easy for me to understand.

☐ The experiment encouraged active participation from the children and sparked their curiosity.

Are there parts of the material that need improvement? If yes, which?

_____

2. Practical use

☐ The material was user-friendly in relation to the children's development and learning.

☐ I needed to adapt the material for it to work with the children's group.

3. Learning and exploration

☐ The children showed interest in the experiment.

☐ The activity supported the children's critical thinking, exploration, and learning in STEM (Science, Technology, Engineering, Mathematics).

4. Personal reflection

☐ I have learned more about leading experiments related to STEM.

☐ I feel more confident working with sustainable learning in the children's group.

Which skills have you developed by conducting the experiment?

_____

Do you have any additional feedback to help improve the material?

_____

Appendix 2: Partner guidelines and data collection materials

This document serves as a guideline for partner organisations. It outlines responsibilities across five areas: conducting teacher training, planning activities with children, supporting individual teacher reflection, compiling data, and conducting focus group interviews.

Part 1: Guide for trained facilitators

Partners are responsible for training at least 20 preschool teachers, either independently or in collaboration with the associated preschools. The training follows the same five-step structure described in Appendix 1. Project goals:

- Goal 1: Train at least 20 preschool teachers or childcare workers from participating preschools in planning and conducting Superhero Lab experiments.
- Goal 2: Ensure at least 40 children aged 4-6 participate in at least two Superhero Lab experiments led by trained teachers.
- Goal 3: Conduct and compile two focus group interviews with different preschool teachers (5-7 participants per group from the participating preschools).

Part 2: Activity plan with the children

Partners support preschools in implementing the Superhero Lab experiments with children. After each experiment, children reflect using the self-assessment form (A2.1 and A2.2).

Part 3: Individual reflection for preschool teachers

Preschool teachers reflect individually after each experiment using A2.3. Partners collect and compile these reflections.

Part 4: Collecting and compiling data

Partners collect A2.2 and A2.3 from the participating preschools and compile the results in A2.4. Compiled results should be submitted to K&R Education (annika@kreducation.se and marie@kreducation.se) on four occasions: 10 October 2025, 15 December 2025, 13 February 2026, and 31 March 2026.

Part 5: Focus group interviews with preschool teachers

Two focus group interviews are conducted with preschool teachers who have participated in the training and worked with the Superhero experiments. Each focus group should consist of 5-7 participants. It is recommended that two people conduct the interview: one to lead the discussion and one to document. Use A2.5. Results should be submitted to K&R Education by 15 December 2025 (first round) and 31 March 2026 (second round).

A2.1: Self-assessment questions for children

Ask children one question at a time and have them point to the face that best represents how they feel.

- How did it feel to do the experiment?
- I thought it was fun to do the experiments together with the others.
- We helped each other in the group when we did the experiment.
- Did the experiment turn out the way you expected?
- Did you learn anything about nature during the experiment?
- I want to do more experiments.

A2.2: Data collection form: children's responses

Fill in how many children gave each answer per question or statement.

Superhero Lab experiment conducted: _____

Number of children who participated: _____

Country: _____

Table A2.1

Children's self-assessment response form (A2.2)

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?			
I thought it was fun to do the experiments together with the others.			
We helped each other in the group when we did the experiment.			
Did the experiment turn out the way you expected?			
Did you learn anything about nature during the experiment?			
I want to do more experiments.			

A2.3: Individual reflection form for preschool teachers

Reflect and respond individually. Tick the boxes if the statements apply.

Experiment conducted: _____

Number of children who participated: _____

Country: _____

1. Design and materials

☐ The material was clear and easy for me to understand.

☐ The experiment encouraged active participation from the children and sparked their curiosity.

Are there parts of the material that need improvement? If yes, which?

_____

2. Practical use

☐ The material was user-friendly in relation to the children's development and learning.

☐ I needed to adapt the material for it to work with the children's group.

3. Learning and exploration

☐ The children showed interest in the experiment.

☐ The activity supported the children's critical thinking, exploration, and learning in STEM.

4. Personal reflection

☐ I have learned more about leading experiments related to STEM.

☐ I feel more confident working with sustainable learning in the children's group.

Which skills have you developed by conducting the experiment?

_____

Do you have any additional feedback to help improve the material?

_____

A2.4: Compiled results template

Goal 1

Country: _____

Number of unique preschool teachers trained to date: _____

Superhero Lab experiments conducted: _____

Goal 2

Number of unique children who participated: _____

Cumulative children's self-assessment totals (from all A2.2 forms):

Table A2.2

Cumulative children's self-assessment results (A2.4)

Question / Statement	Happy face	Neutral face	Sad face
How did it feel to do the experiment?			
I thought it was fun to do the experiments together with the others.			
We helped each other in the group when we did the experiment.			
Did the experiment turn out the way you expected?			
Did you learn anything about nature during the experiment?			
I want to do more experiments.			

Number of unique teachers who completed the individual reflection (A2.3):

_____

A. Design and materials

☐ The material was clear and easy for me to understand. Total ticks: _____

☐ The experiment encouraged active participation from the children and sparked their curiosity. Total ticks: _____

Open responses — Are there parts of the material that need improvement?

_____

B. Practical use

☐ The material was user-friendly in relation to the children's development and learning. Total ticks: _____

☐ I needed to adapt the material for it to work with the children's group. Total ticks: _____

C. Learning and exploration

☐ The children showed interest in the experiment. Total ticks: _____

☐ The activity supported the children's critical thinking, exploration, and learning in STEM. Total ticks: _____

D. Personal reflection

☐ I have learned more about leading experiments related to STEM. Total ticks: _____

☐ I feel more confident working with sustainable learning in the children's group. Total ticks: _____

Open responses — Which skills were developed? Additional feedback?

_____

A2.5: Focus group interview questions

Country: _____

Design and materials

- How clear is it to understand how the experiments should be used?
- Is the material easy to adapt and modify based on your children's group's conditions and needs? If yes, how? If no, why not?
- Are there parts of the material that need development? If yes, which parts, and how should they be developed?

Children's development and learning

- How do you feel the experiments engage and spark the children's curiosity?
- In what ways do the children get to be active and explore during the experiments?
- How do the experiments help the children develop their critical thinking?
- What have you noticed the children have learned or developed through doing the experiments together, linked to the curriculum?

Teachers' learning

- How confident do you feel about what GreenComp is? Rate yourself on a scale from 1-5, where 1 is not at all and 5 is to the highest degree.
- How has your work with Superhero Lab influenced your understanding of sustainability and STEM?
- In what way has the training changed how you work with practical experiments and scientific exploration?
- How have your self-confidence and motivation been affected by using the Superhero Lab experiments?
- How do you perceive your confidence in working with sustainable learning in the children's group now compared to before?

Other

Do you have anything else to add?

Compile responses and submit to annika@kreducation.se and marie@kreducation.se.

Timeline: WP4 for all partners

Table A2.3

Implementation timeline for partner countries

Period	Activities and deadlines
June 2025	Teacher training in Solkan. Each partner creates a plan for teacher training during Autumn 2025 and Spring 2026.
July–August 2025	Vacation and preparation time.
September 2025	Partners begin implementing teacher training in the preschools based on the partner guidelines (Steps 1-5).
October 2025	Deadline 1 October: A2.2 and A2.3 submitted to the responsible trainer at each preschool, then to the partner organisation. Deadline 10 October: A2.4 compiled results submitted to K&R Education.
November 2025	First round of focus group interviews with preschool teachers. Each group: 5-7 participants. Use A2.5.
December 2025	Deadline 1 December: A2.2 and A2.3 submitted. Deadline 15 December: A2.4 compiled results submitted to K&R Education. Deadline 15 December: First focus group results (A2.5) submitted to K&R Education.
January 2026	K&R Education compiles and analyses the collected information.
February 2026	Deadline 2 February: A2.2 and A2.3 submitted. Deadline 13 February: A2.4 compiled results submitted to K&R Education.
March 2026	Deadline 20 March: A2.2 and A2.3 submitted. Deadline 31 March: A2.4 compiled results and second focus group results (A2.5) submitted to K&R Education.
April 2026	K&R Education compiles and analyses all collected information.
May 2026	Deadline for completion of WP4: 31 May 2026.
June 2026	K&R Education presents the results of WP4.