

# Methods

In this study, I analysed how an innovative ESD intervention influenced students' sustainability attitudes, sustainability behaviours and efficacy beliefs as components of sustainability competencies over time. To quantitatively test the effect of the ESD intervention, I assessed long-term effects by measuring data at three different measurement points (pre-intervention, immediately post-intervention and one year post-intervention) using the same self-reported survey on sustainability behaviours and sustainability attitudes. I also complemented the data collection with a survey on efficacy beliefs at the third measurement point. I used the data collected by Pauli (2023) for measurement points one and two. In total, I analysed 206 self-reported surveys.

## Project description KlimaRatSchule

To analyse the changes in sustainability competencies over time, I chose the ESD intervention KlimaRatSchule (KRS). The KRS project aims to promote and establish an active climate protection culture in schools through democratic participation processes. Students investigate the carbon footprint of their own school and carry out a democratic participation process, the results of which are used to develop the school's own climate protection concept (KRS website, 2024). The project was piloted in two schools in Freiburg in January 2023. The project was designed by Solare Zukunft e.V., IZT, Energie- und Umweltzentrum am Deister e.V. and Green City e.V.. The project can be divided into 3 phases: in the first phase, a group of volunteers (involved group) was formed and collected data on the school's emissions. In the second phase, a micro-citizen report for schools was conducted with randomly selected students from the school, and in the third phase, the project partners developed a roadmap with concrete goals and steps for the school to reduce its carbon footprint based on the students' work in phases one and two (KRS website, 2024). Although the project included the very interesting aspect of participation, unfortunately, due to severe data limitations, the group that was part of the micro-citizen report had to be excluded from the analysis and therefore the analysis of the participatory aspect of the project also had to be excluded. The ESD intervention aimed to achieve a real impact in terms of CO<sub>2</sub> emission reduction, while at the same time achieving climate and democratic education.

The ESD intervention can be classified as follows in terms of learning and teaching methods and procedures. The involved group, which was voluntarily participating, assessed different dimensions of the school's CO<sub>2</sub> emissions in a self-regulated and self-directed manner. They applied their knowledge and problem-solving skills in a real-world, group-based, solution-oriented environment (KRS website, 2024). Considering all these learning methods and procedures, this ESD intervention can be considered as an innovative ESD intervention (as opposed to more traditional teaching-learning formats) (Riess et al., 2022).

## Study design and data collection

The research was designed as an empirical, long-term, quasi-experimental study. To collect the data, a digital survey was created that mapped the constructs of the Theory of Planned Behaviour, which was the same as that used by Pauli (2023) for the first two measurement points. Accordingly, attitudes, subjective norms, perceived behavioural control, intentions and behaviours with regard to climate protection were measured. In addition, the survey was complemented by a scale measuring efficacy construct based on the Triple A framework. Accordingly, personal and collective efficacy beliefs as well as action and aim-focused efficacy beliefs were assessed.

The survey was conducted at a secondary school in the Freiburg area, the ANGELL school, a Montessori centre in the city, which is state-recognised and run as an independent non-profit organisation. Following a quasi-experimental design, two comparison groups were formed for the study. The involved group was part of the voluntary KRS expert group that carried out the CO<sub>2</sub> measurements in the school, and the control group did not participate at all. The surveyed students were classified into one of the groups by being asked in the survey whether they participated in the KRS school group. Both groups were surveyed at three different

measurement points related to the progress of the intervention, see Table XX. At all three data collection points, the same self-report survey was administered, with the idea of interviewing the same students at three different points in time. This combination of a longitudinal study and a treatment group-control group comparison promises to be particularly informative about the relationship between project participation and impact (Ssossé et al., 2021).

There were some limitations in collecting data at the third measurement point. Due to government regulations, the survey could only be conducted in one of the schools participating in the KRS project. Conducting the survey in more than one school outside the same school authority area would have required permission from the school supervisory authority (VerbundFDB, 2022), which was not possible due to time constraints. In the school where the survey was carried out, it was important to minimise disruption to the students. Given the challenge of trying to capture the same students as for the previous two data points, it was decided to survey students in grades 10 and 11. At the first two measurement points, at least 85% of the students were in Years 9 and 10 (Pauli, 2023), thus maximising the likelihood of interviewing the same students a year later, see Table XX for overview of data collection.

Table 1: Overview data collection at the school at three different measurement points.

# Measurement points	Intervention phase	Date of data collection	Classes surveyed	Data collector	N data collected (control group/involved group)
Measurement Point 1 (MP1)	Pre-intervention (Project start)	09.02.23	7-11	Data collected by Pauli	76 (56/20)
Measurement Point 2 (MP2)	Immediately post-intervention (end of active phase of project)	10.05.23	7-11	Data collected by Pauli	81 (63/18)
Measurement Point 3 (MP3)	One year post-intervention	13.05.24 – 06.06.24	10-11	Data collected by Gargya	49 (42/7)

## Survey tools

All information below refers to the Measurement Point 3. The survey was conducted in German, see the German questions used in Appendix A. The items of a scale were asked together in one block. The items were all in the form of statements to be answered on a 4-point Likert scale (0 = strongly disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = strongly agree). An ‘I don’t know/No response’ option was provided for each item. The completed survey was entered into the Sco-Scie online survey application and completed online by the students. The suggested test time was 10-15 minutes. In accordance with the state’s educational research and data policies, parental and school principal consent was obtained prior to the assessment (see Appendix B). Participation was voluntary, i.e. students received no credit or financial reward. Participants were assured of complete confidentiality and anonymity. Unfortunately, the researcher could not be present at the third measurement point. Therefore, a slide was prepared with the essential information (see Appendix C).

### *Sustainability attitudes and sustainability behaviours*

In order to ensure the comparability of the data with the first two measurement points and to allow for long-term assessment, the exact same questions were used at the third measurement point, based on Pauli (2023). The scale was designed on the basis of the Theory of Planned Behaviour and included elements of attitudes, subjective norms, perceived behavioural control, intentions and behaviour, see Table 2. Pauli (2023) collected the questions from various literature sources and partly adapted them to a high school context.

Table 2: The scales used to measure sustainability attitudes and sustainability behaviours based on the Theory of Planned Behaviour. Attitudes, subjective norms, perceived behavioural control and intentions were combined to form sustainability attitudes, and self-reported behaviours were taken as sustainability behaviours. The scales have been copied one to one from Pauli (2023). The questions here have been translated to English for consistency, see German version in Appendix A.

#	Scale	Item	Theoretical classification	Question (drawn from original sources in English)	Source
1	Attitudes towards climate-protecting behaviour	AT1	Sustainability attitude	The environment in Germany is in danger because of global climate change.	Masud et al. (2016)
2		AT2		Current global warming is a natural not manmade phenomenon.	
3		AT3		Climate change damages the natural environment and wildlife in Germany.	
4		AT4		I'm willing to pay a certain amount to reduce the im-pact of climate change.	
5	Subjective Norms	SN1		My family often discusses climate change or global warming.	Lin (2013)
6		SN2		My peers often discuss climate change or global warming.	
7		SN3		My classmates might criticize me if I don't take action to protect the climate.	
8	Perceived Behavioural Control	PBC1		I believe I can contribute to mitigating the effects of climate change.	Pouya und Niyaz (2022)
9		PBC2		I can help reduce carbon emissions through the actions I take in my daily life.	
10	Intentions	IN1		It is my responsibility to encourage my neighbours to notice climate change.	Pouya und Niyaz (2022)
11		IN2		I am willing to adopt and apply eco-friendly practices in my daily life.	
12		IN3		I am ready to do anything to reduce the impact of climate change.	
13	Behaviours	B01	Sustainability behaviour	I have reduced my intake of meat over the last month.	Lin (2013)
14		B02		I buy fruit produced in Germany and avoid buying imported fruit (e.g. bananas, kiwis).	
15		B03		I often store food in my refrigerator that is past its use-by date.	
16		B04		I am more concerned about price when I purchase electrical appliances.	
17		B05		I purchase electrical appliances that have energy saving labels.	
18		B06		I turn lights and water taps off as much as possible.	
19		B07		I unplug appliances that are temporarily not in use.	
20		B08		I mainly drive or am driven by a car or a motor-scooter.	
21		B09		I use elevators and rarely use stairs.	
22		B10		I support an increase of fuel tax to reduce the use of fossil fuels.	

Efficacy beliefs

The efficacy scales were based on Hamann et al.'s (2024) questions. Further adaptations for operationalisation in this context were based on their recommendations. For all questions, I included the marker words 'if I/we want to' to avoid the risk of capturing the desirability of the aim instead of their efficacy beliefs. I considered two agents as social identities central to the context of the study. In addition to the individual, I also considered 'we as students' to be a relevant social identity in this context, given the nature of the project to shape the school environment, their primary occupation as students, and the educational goal of the intervention. In order to be able to compare personal and collective efficacy beliefs, I chose three questions and repeated them with the same wording, except that I changed the agent from "I" to "we, as students". In the questions, I included both action and aim-focussed links (see Table 3). I partially adapted some of the questions to the school context by specifying decision-makers or organisations as "school principals or schools" (for the original questions, see Hamann et al. (2024)). Due to time constraints, I could not conduct pilot studies to test the desirability of actions and aims, so I kept the questions as suggested by Hamann et al. (2024), see Table 3 for all questions.

Table 3: The scales used for efficacy beliefs were based on the Triple-A framework (Agent-Action-Aim) by Hamann et al. (2024). The scales allowed differentiation between personal and collective efficacy beliefs and between aim and action focus. The questions here have been translated to English for consistency, see German version in Appendix A.

#	Scale	Item	Theoretical classification question	Question	Source
1	Personal efficacy beliefs	SW01_01	Aim-focussed	I believe that my own actions can make a contribution to climate protection if I want to.	Hamann et al. (2024)
2		SW01_02	Aim-focussed	I believe that I can promote climate protection by educating people around me about climate change if I want to.	
3		SW01_03	Action-focussed	I don't think I'm in a position to stand up for climate protection.	
4		SW01_04	Aim-focussed	I believe that I am in a position to convince others to do more for climate protection if I want to.	
5		SW01_05	Aim-focussed	I don't believe that I have any way of influencing climate change.	
6		SW01_06	Action-focussed	I believe that I can influence how my school directors or my school acts in relation to climate protection if I want to.	
7		SW01_07	Aim-focussed	I believe that I can support my school directors or school in campaigning for climate protection if I want to.	
8		SW01_08	Aim-focussed	I believe that I can make a meaningful contribution to climate protection in collaboration with others if I want to.	
9	Collective efficacy beliefs	CS01_01	Aim-focussed	We, as students, can make a contribution to climate protection through our actions if we want to.	
10		CS01_02	Action-focussed	We, as students, can influence how our school directors or school acts in relation to climate protection if we want to.	
11		CS01_03	Aim-focussed	We, as students, are in a position to convince others to do more for climate protection if we want to.	

## Data processing

I conducted all data processing and analyses in R v. 4.0.2. (R Core Team, 2021).

Data cleaning and transformation involved several steps, see also the R code in Appendix E. As my analysis

was based on (cleaned and transformed) data from Pauli (2023), I carried out very similar steps for my data. I excluded incomplete surveys. I also excluded surveys with more than 25% ‘I don’t know/No answer’ responses. With a total of 32 items, this meant deleting cases with 8 or more unspecified responses. This ensured that each scale could be calculated from as many items as possible. According to Döring and Bortz (2016), individual missing values in the dataset are not a problem for further analysis. To avoid collecting data where students simply clicked through the survey as quickly as possible without answering the questions conscientiously, I excluded surveys with a timestamp of less than 120 seconds. Likert scale responses were coded into numerical values for each item: 0 = strongly disagree, 1 = somewhat disagree, 2 = somewhat agree, 3 = strongly agree, so that in the end a low score indicated a low level of SA/SB/efficacy beliefs, and a high score indicated a correspondingly high level. To ensure that this was the case for all questions, items AT2, B03, B04, B08, B09, SW01\_03, SW01\_05 had to be inverted, as they contained negative formulations. The measurement level of the Likert-scale items was defined as metric, which allowed the calculation of mean values. After reviewing the quality criteria for quantitative research, described below, the mean of the scales, reflecting the five components of the TPB and personal and collective efficacy beliefs, were calculated from the individual items.

## Data analysis

In order to quantify the relationship between the components of sustainability competencies across different levels of involvement and over time, I used different methods of statistical analysis. In the area of inductive statistics, methods from dependency analysis were used. As this is an explanatory study, the hypothesis test with classical significance test was applied for the inferential statistical evaluation (Döring & Bortz, 2016). The significance level  $\alpha = 0.05$  was used for the significance test. If a p-value was below this threshold, it was marked with one asterisk as ‘significant’, with two asterisks as ‘very significant’ or with three asterisks as ‘highly significant’, see Table 4.

Table 4: Overview of p-value labels in the significance tests.

Labelling	Designation	Threshold values
*	significant	0.05
**	very significant	0.01
***	highly significant	0.001

In order to choose the appropriate statistical analysis methods, I first tested the distribution of my data. This was done using the Shapiro-Wilk test, whose null hypothesis is that there is a normal distribution. Almost all scales were below the significance threshold of 0.05. This means that the null hypothesis had to be rejected for most scales. The data were therefore largely free of distribution. In order to ensure uniformity and thus comparability, non-parametric tests were used for all scales despite metric scaling, as required by a distribution-free data base (University of Zürich, 2024).

In order to answer some of my research questions (RQ1 and RQ3) by comparing the scale means between groups, I conducted the Wilcoxon test. This test is suitable for distribution-free, interval-scaled dependent samples and determines whether the central tendencies of two dependent samples are different (University of Zürich, 2024). Given the lack of normal distribution in my data and the need to compare tendencies between groups, I considered the Wilcoxon test appropriate to provide statistical answers to the question of whether the central tendencies of the groups differ. This approach was applied to all scales relevant to my research questions: sustainability attitudes (comprising scales of attitude, intention, perceived behavioural control and subjective norms), sustainability behaviours (all based on the Theory of Planned Behaviour), collective and personal efficacy beliefs, and action and aim-focused efficacy beliefs.

To answer my research question (RQ1) on whether sustainability attitudes and sustainability behaviours differ over time, I conducted the Kruskal-Wallis test. This test is suitable for distribution-free, ordinal-scaled independent samples and determines whether the central tendencies of more than two independent samples are different (University of Zürich, 2024). Given the lack of normal distribution in my data and the

need to compare three different measurement points for each SA and SB in each group, I considered this test sufficient to provide statistical answers. I decided not to perform any post-hoc tests as the Kruskal-Wallis test showed no statistical differences between the groups.

To answer my research question (RQ2) regarding the relationship between SA/SB and efficacy beliefs, I conducted the Spearman rank correlation test. This test is suitable for distribution-free, ordinal-scaled samples and calculates the linear correlation between two variables (University of Zürich, 2024). The rank correlation can range from -1 to +1, where -1 indicates a perfect negative correlation and +1 indicates a perfect positive correlation. If  $r_s=0$ , there is no correlation.

## Quality Criteria testing

In the case of a fully standardised questionnaire survey, it must be subjected to a number of checks to ensure the reliable quality of the results. In order to draw conclusions about the population in the sense of inductive statistics, it must be ensured that the sample is representative of the population. This is an aspect of external validity (Döring & Bortz, 2016). The KRS school group met at the school itself and thus became the involved group in the study, consisting of about 20 students, which fluctuated somewhat during the school year. All the classmates of the students in the KRS group were defined as the basic population of the control group, as they were exposed to similar other conditions and thus comparability was ensured. No information was available on the exact size of this population, so an estimate is given in Table XX. The sample from this control group was determined by the willingness of the students approached, to participate. Calculations of ideal sample sizes, based on a 90% confidence interval and a 10% margin of error (Qualtrics, 2024), show that for measurement point 3, neither the control group nor the involved group were representative of the population, which limits the conclusions that can be drawn about the population, see Table 5.

Table 5: Comparison of ideal and actual sample sizes in relation to the population.

	<b>Measurement point 1</b>		<b>Measurement point 2</b>		<b>Measurement point 3</b>	
<b>Group</b>	Control group	Involved group	Control group	Involved group	Control group	Involved group
<b>Total population</b>	225	20	225	20	150	20
<b>Ideal N (90% confidence interval + 10% margin of error)</b>	> 53	> 16	> 53	> 16	> 47	> 16
<b>Actual N</b>	56	20	63	18	<b>42</b>	<b>7</b>

For reliability analysis, Cronbach's alpha, the most common measure of scale reliability (Field & Miles, 2010), was calculated to assess the internal consistency of the subscales. Reliabilities for measurement points 1 and 2 were calculated and confirmed by Pauli (2023). The results for measurement point 3 are shown in Table 6 below.

Table 6: Reliabilities (Cronbach's alpha) for sustainability competency scales at measurement point 3.



Scale	Number of items	Cronbach's $\alpha$
Attitude	4	0.65
Subjective Norms	3	<b>0.41</b>
Perceived behaviour control	2	0.77
Intention	3	0.86
Behaviour <sub>1</sub>	8	0.78
Personal efficacy beliefs	8	0.90
Collective efficacy beliefs	3	0.90

<sup>1</sup> To calculate the Cronbach's  $\alpha$  the variables B03 and B04 were omitted, because there were not considered fitting and essential. The content validity should not be jeopardised by this.

In the literature, Cronbach's alpha values above 0.6 are considered good for short scales (above 6 items). For longer scales (8 items), values above 0.7 can be considered good (Krüger et al., 2014). Almost all scales achieved satisfactory results in this test, with the exception of the subjective norms scale. This lower reliability should be taken into account in further analyses, although the literature also points out that these thresholds should not be given too much significance (Schmitt, 1996).

The full R script and the link to my GitHub repository can be found in the appendices.