# Methods

In this study, I analysed how an innovative ESD intervention influenced students’ sustainability attitude, sustainability behaviour and self-efficacy beliefs as components of sustainability competencies over time. To quantitatively test the effect of the ESD intervention, I assessed long-term effects by using data measurements at three different measurement points (before, straight after intervention, and one year after intervention), using the same self-reported survey on sustainability behaviour and sustainability attitude. Furthermore, I complemented data collection with a survey on self-efficacy at the last measurement point. I used the data collected at the measurement point one and two from Pauli (2023). In total I analysed 206 self-reported surveys.

## Project description KlimaRatSchule

For analysing the changes of 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. The students deal with 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 Projektbeschreibung, 2024). The project was piloted at 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) formed and collected data on the school’s emissions. In a second phase, a micro citizen report for schools was conducted with randomly drawn students from the school and in the third phase the project partners developed a roadmap with concrete aims and steps for the school to reduce their carbon footprint based on the work of the students in phase one and two. Although the project included the very interesting aspect of participation, unfortunately due to severe data limitations, the group, which were part of micro citizen report, had to be excluded from the analysis and thereby also the analysis of the participative aspect of the project. The ESD intervention aimed to achieve a real impact of CO2 emissions reductions while at the same time achieving climate and democratic education.

The ESD intervention can be classified as following regarding the learning and teaching methods and procedures. The involved group, which found themselves to be voluntarily participating, assessed different dimensions of the CO2 emissions of the school in a self-regulated and self-directed manner. Applying their knowledge and problem-solving skills in real world, group-work based, solution-oriented setting (REF KRS). Given all these learning methods and procedures, this ESD intervention can be considered as an innovative ESD intervention (in contrast to more traditional teaching-learning formats) (REF MONI).

## Study design and data collection

The research was designed to be an empirical, long-term, quasi-experimental study. To collect the data, a digital survey was compiled that mapped the constructs of the Theory of Planned Behaviour, which was the same one Pauli (2023) used for the first two measurement points. Accordingly, attitudes, subjective norms, perceived behavioural control, intention and behaviour with regard to climate protection were surveyed. Additionally, the survey was complemented by a scale that mapped the constructs of self-efficacy based on the triple A framework. Accordingly, individual and collective self-efficacy beliefs, and action- and aim- focussed self-efficacy beliefs were surveyed.

The survey was conducted at a secondary school in the Freiburg area, namely the ANGELL school, which is 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 survey. The involved group was part of the voluntary KRS-expert group, conducting the CO2 measurements at the school and the control group did not participate at all. They were categorised into one of the groups by being asked in the survey whether they took part in the KRS school group. Both groups were surveyed at three different measurement points related to the intervention progress, see Table XX. At all three data collection points, the same self-reported survey was conducted with the idea of surveying the same students at three different time points. *This combination of longitudinal study and treatment group-control group comparison promises to provide particular informative value about the relationship between project participation and impact (N. Döring & Bortz, 2016, p. 1023; Ssossé et al., 2021, p. 9). (*The involvement of the three different groups of involvement was not possible due to too small sample sizes of group one in the data collection. I then excluded this group from all analysis.)

Conducting data at the third measurement point was subject to some limitations. Due to state regulations, the survey was only carried out at one of the participating schools of the KRS project. Conducting the survey at more than one school outside the same territory of a school authority would have needed a permit from the upper school supervisory authority (REF VERBINF FDB), which was not possible to be obtained due to time limitations. At the school where the survey was conducted, the disruption of the students had to be kept to a minimum. Given the challenge of trying to capture the same students as for the previous two data points, it was decided to survey the students of the grades 10 and 11. At the first two measurement points, at least 85% of the students were in the grades 9 and 10 (REF PAULI), therefore maximising the likelihood of surveying the same students again one year later.

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **# Measurement points** | **Intervention phase** | **Data of data collection** | **Classes surveyed** | **Data collector** | **Presence of investigator** | **N data collected (control group/involved group)** |
| First measurement point | Project start | 09.02.2023 | 7-11 | Data collected by Pauli | Yes | 76 (56/20) |
| Second measurement point | End of the active phase of the project | 10.05.2023 | 7-11 | Data collected by Pauli | Yes | 81 (63/18) |
| Third measurement point | Approximately one year after the intervention | 13.05.2024 – 06.06.2024 | 10-11 | Data collected by Gargya | No | 49 (42/7) |

### Survey tools

All the following information is regarding the third measurement point. The survey was conducted in German, see the German questions used in Appendix XX. The items of a scale were asked together in a block. The items were all in the form of statements and were 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 given for each item. The completed survey was entered into the online survey application Sco-Scie and was answered online by the students. Proposed testing time was 10-15 minutes. Following the state’s educational research and data guidelines, parental and the school principals’ consent were obtained prior to the assessment (see Appendix XX). Participation was voluntary, i.e., students did not get any credit or monetary reward. Participants were assured of full confidentiality and anonymity. Unfortunately, at the third measurement point, the investigator could not be present herself. Therefore, a slide was prepared with the essential information (see Appendix XX).

#### Sustainability attitude and sustainability behaviour

To ensure comparability of data to the first two measurement points and allow for long term assessment, at the third measurement point, the exact same questions were used, based on Pauli (2023). The scale was designed based on the theory of planned behaviour and included elements of attitudes, subjective norms, perceived behaviour control, intentions, and behaviour, see Table XX. Pauli (2023) gathered the questions from different literature sources and partly adjusted them to fit a high school context.

Table XX: The scales used regarding the sustainability attitude and sustainability behaviour based on the theory of planned behaviour. Attitude, subjective norms, perceived behaviour control and intentions combine themselves to sustainability attitude; and self-reported behaviours to sustainability behaviour. The scales were copied one-to-one from Pauli (2023).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | 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. |

#### Self-efficacy beliefs

The scales for the self-efficacy beliefs were based on the questions by Hamann et al. (2024). Further adjustments to adapt to this context for operationalisations 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 self-efficacy beliefs. I considered two agents, as social identities, to be central to the context of the study. Next to the individual, I also considered “we as students” to be a relevant social identity in this context, given the nature of the project about shaping the school environment, their main occupation being a student, as well as fitting to the educational goal of the intervention. To be able to compare individual and collective self-efficacy beliefs, I picked three questions and repeated them with the same wording, except changing the agent from “I” to “we, as students”. In the questions, I included both action- and aim-focused links (see Table XX). I partly adjusted some of the questions to the school context, by specifying decision-makers or organisations to be “school directors or schools” (for original questions, see Hamann et al. (2024)). Due to time limitations, I could not conduct pilot studies to test for desirability of actions and aims, so I kept the questions as they were suggested by Hamann et al. (2024), see Table XX for all questions.

Table XX: The scales used regarding the self-efficacy beliefs based on the Triple-A framework (Agent-Action-Aim) by Hamann et al. (2024). The scales allow for differentiation between individual and collective self-efficacy beliefs and between aim- and action-focus.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Scale** | **Item** | **Theoretical classification question** | **Question** | **Source** |
| 1 | Individual self-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. |
| 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 self-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. (REF R Core Team, 2021).

The data cleaning and data transformation involved several steps, see also Appendix XX. Given that my analysis was based on (cleaned and transformed) data by Pauli (2023), I conducted very similar steps for my data. I excluded incomplete surveys. I also excluded surveys with more than 25% answers of “I don’t know/ No response”. Given 32 items in total, that meant deleting cases with the unspecified answer of 8 or more times. 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 data set are not a problem for further analysis. To avoid capturing data were students just clicked through the survey as fast as possible without answering the questions conscientiously, I excluded survey with a timestamp of less than 120 seconds. The answers on the Likert scale 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 value indicated a low level of SA/SB/ self-efficacy beliefs, and a high value indicated a high level accordingly. To ensure that this was the case for all questions, the items AT2, B03, B04, B08, B09, SW01\_03, SW01\_05 had to be inverted, as negative formulations were present here. Following Döring and Bortz (2016, p. 269), the measurement level of the Likert-scaled items was defined as metric, which enabled the calculation of mean values. After checking for the quality criteria of quantitative research, which is described in the following, the scales were calculated from the individual items, which reflect the five components of the TPB and the individual and collective self-efficacy beliefs. For all scales the mean values of the associated items were calculated.

## Data analysis

To quantify the relationship of components of sustainability competencies across level of involvement and over time, I used different statistical analysis methods. In the area of inductive statistics, methods from dependency analysis were applied. This being an explanatory study, for the inferential statistical evaluation the hypothesis test with classical significance test was applied (N. Döring & Bortz, 2016). The significance level α = 0.05 was used for the significance test (Döring und Bortz). If a p-value was below this threshold, this was labelled with one asterisk as ‘significant’, with two asterisks as ‘very significant’, or with three asterisks as ‘highly significant’ (cf. Krüger et al., 2014, p. 291), see Table XX.

Table XX: Overview labelling of the p-values in the significance tests

|  |  |  |
| --- | --- | --- |
| **Labelling** | **Designation** | **Threshold values** |
| \* | significant | 0.05 |
| \*\* | very significant | 0.01 |
| \*\*\* | highly significant | 0.001 |

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

To answer some of my research questions (RQ1a and RQ2) 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 (Uni Zürich, 2024). Given the lack of normal distribution in my data and the need to compare tendencies between groups, I deemed the Wilcoxon test appropriate for providing statistical answers regarding whether the central tendencies of the groups differ. This approach was applied to all scales relevant to my research questions: sustainability attitude (comprising scales of attitude, intention, perceived behavior control, and subjective norms), sustainability behavior (all based on the Theory of Planned Behavior), collective and individual self-efficacy beliefs, and action- and aim-focused self-efficacy beliefs.

To answer my research question (RQ1a) regarding whether sustainability attitudes (SA) and sustainability behaviors (SB) differed 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 (Uni Zürich). Given the lack of normal distribution in my data and the need to compare three different time points for each SA and SB in each group, I deemed the test sufficient to provide statistical answers. I decided not to conduct any post-hoc tests, as the Kruskal-Wallis test showed no statistical differences between the groups.

To answer my research question (RQ1b) regarding the relationship between SA/SB and self-efficacy beliefs, I conducted the Spearman rank correlation test. This test is appropriate for distribution-free, ordinal-scaled samples and calculates the linear correlation between two variables (Uni Zürich). The rank correlation can range from -1 to +1, with -1 indicating a perfectly negative correlation and +1 indicating a perfectly positive correlation. If rs=0r\_s = 0rs​=0, there is no correlation (Tachtsoglou & König, 2017).

## Quality Criteria testing

In the case of a fully standardised questionnaire survey, this must be subjected to a number of checks in order to ensure the reliable quality of the results. To be able 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 (N. Döring & Bortz, 2016, p. 104). The KRS school groups came together at the school itself and thus became the involved group in the study, consisting of around 20 students, which fluctuated somewhat over the course of the school year. All classmates of the students in the KRS school group were defined as the basic population of the control group, as they are exposed to similar other conditions and thus comparability is guaranteed. No information is available on the exact size of this population, which is why an estimated value was entered in Table XX. The sample from this control group was determined by the willingness of the students approached to participate. Calculations of the ideal sample sizes based on 90% confidence interval and 10% margin of error (REF Qualtrics) show that for measurement point 3, neither for the control group, nor for the involved group, representative amounts were drawn from the population, allowing limited conclusions about the population.

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

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

For reliability analysis, Cronbach’s α, which is the most common measure of scale reliability (Field, 2012 REF EWM DEV), was calculated to assess the internal consistency of the subscales. Reliability for measurement points 1 and 2 were calculated and confirmed by Pauli (2023). The results for measurement point 3 can be seen in the following Table XX.

Table XX: Reliabilities (Cronbach’s α) for sustainability competency scales at measurement point 3.

|  |  |  |
| --- | --- | --- |
| **Scale** | **Number of items** | **Cronbach’s α** |
| Attitude | 4 | 0.65 |
| Subjective Norms | 3 | **0.41** |
| Perceived behaviour control | 2 | 0.77 |
| Intention | 3 | 0.86 |
| Behaviour1 | 8 | 0.78 |
| Individual self-efficacy beliefs | 8 | 0.90 |
| Collective self-efficacy beliefs | 3 | 0.90 |

1 To calculate the cronbach’s α 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 literature, values of over 0.6 for Cronbach's alpha are considered good for short scales (≤ 6 items) (Krüger et al., 2014). For longer scales (8 items), values from 0.7 can be considered good (N. Schmitt, 1996). Almost all of the scales achieved satisfactory results in this test, except the subjective norms one. This lower reliability should be taken into account in further analyses, although the literature also points out that these threshold values should not be given too much significance (SchMIDT REF EWM DEV).

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