15 Application-Oriented Development of Outcome Indicators for Measuring Students' Sustainability Competencies: Turning from Input Focus to Outcome Orientation

Eva-Maria Waltner, Anne Overbeck, and Werner Rieß

Abstract

Since the Brundtland Report and the Agenda 21 conference in Rio, many Education for Sustainable Development (ESD) programs have been launched. However, until now, empirical data on the impact and outcome of ESD initiatives within educational settings is scarce. This chapter explores the assessment of sustainability competencies including cognitive, affective, and behavioral domains, by presenting different possibilities, results, and limitations of ESD assessment goals and frameworks. This contribution emerges from a collaboration of researchers on the operationalization and measurement of ESD outcomes at the University of Education Freiburg and the Otto-von-Guericke University Magdeburg. Data from a longitudinal measurement with students in secondary schools (grades 5–8, n = 1324, age 9–16) in the state of Baden-Württemberg was analyzed to gain a clearer picture of the development of students' sustainability competencies within 1 school year. This data shows that measuring the outcome of ESD teaching programs is possible. Using these empirical measures could thereby facilitate decision-making on ESD measures for many different levels.

Keywords

ESD · Measurement · Outcome · Sustainability competencies · Indicators

15.1 Introduction

Despite long and ongoing attempts to change toward a more sustainable society, environmental problems remain or have even worsened in many ways (e.g., Cai et al., 2016; Lenton et al., 2008; Schellnhuber, 2006). Based on these observations, a growing interest has emerged which calls for a shift from the Input focus of indicators (e.g., policies such as the international treaties on climate change or educational measures) to an Outcome orientation (e.g., actual changes in emissions or the development of sustainability competencies (SCs) on the students/stakehold-

E.-M. Waltner (✉) · W. Rieß
Department of Biology and Pedagogy of Biology, University of Education Freiburg, Freiburg, Germany
e-mail: eva-maria.waltner@ph-freiburg.de

A. Overbeck
Personality and Social Psychology Division, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

G. Karaarslan-Semiz (ed.), *Education for Sustainable Development in Primary and Secondary Schools*, Sustainable Development Goals Series, https://doi.org/10.1007/978-3-031-09112-4_15

ers' level) (e.g., Burford et al., 2016; Waltner et al., 2018, p. 12). In this analysis, we focus on the school level of Education for Sustainable Development (ESD) measures. Though the students who are taught in our schools right now are not the decision-makers of today, they eventually will be the decision-makers of tomorrow. Some effects of the educational measures (e.g., teaching, whole institutional approach) might only be empirically verifiable in the long term or in general not clearly be attributable to a specific measure, due to the complexity of the interaction of many variables affecting, for example, sustainability awareness. However, from these circumstances it cannot be deduced or assumed that there are no characteristics at all which can be recorded empirically. These considerations show that when shifting the attention from the Input to the Output orientation of ESD measures, we might need more long-term assessments and additional method orientations to evaluate the impact. Until now, empirical data on the long- as well as the short-term impact of ESD initiatives within educational settings is scarce. This chapter explores the assessment possibilities of SCs, including cognitive, affective, and behavioral domains, by presenting different approaches and some exemplary results without the ambition of covering ESD assessment methods or the SCs in their entirety.[1]

Before being able to measure the desired outcomes, we need clarity about the goals which are targeted by ESD measures. In this article, the term ESD is used to describe the sum of all actions by which people seek to promote learners' (*or their own*) SCs, i.e., enabling them to facilitate a sustainable development (Waltner et al., 2019). ESD goal dimensions on the outcome level of learners can be divided into cognitive, affective-motivational, and behavioral facets as well as sub-competencies which cannot be clearly assigned to only one of these three areas. Input policies which want to achieve ESD goals on various levels of concreteness can be found in regional (e.g., educational plans, municipal policy) as well as on the inter- or supranational level (e.g., sustainable development goals (SDGs), Paris Agreement).

After defining the outcome which will be considered as relevant to assessing the success of ESD, the indicators which are developed on this basis have to be tested in extensive studies. This contribution emerges from a collaboration of researchers on the operationalization and measurement of ESD outcomes at the University of Education Freiburg and the Personality and Social Psychology Division of the Otto-von-Guericke University Magdeburg, Germany. The former mainly used the local educational plan as ESD goal orientation, the latter the global approach of the SDGs. The data from this first project, a longitudinal measurement with students in secondary schools (grades 5–8, $n = 1324$, age 9–16) in the state of Baden-Württemberg, will be presented. We will further discuss this approach in the light of generalizability of the measurement approaches and results.

The ultimate normative goal of ESD is the actual development of a more sustainable society through sustainable behavior or at least the facilitation of each individual to become a part of sustainable development – see, e.g., the definition of ESD on the local level: Education for Sustainable Development enables learners to make informed decisions and act responsibly for the protection of the environment and for a functioning economy and a just world society for current and future generations (Ministry of Education Baden-Württemberg, 2016, *translated by the authors*). This definition resembles many other internationally used definitions (e.g., Rieckmann, 2018; UNESCO, 2017). Consequently, good outcome measurement needs validity criteria to ensure that these gauges can assess the achievement of these goals – in our case, the competencies which enable the learners (among other things) to contribute to real-world change. To demonstrate how this can be achieved exemplarily, we will present two findings which emerged from the work on this data. First, we performed an ad hoc scale validation, using items from the newly developed ESD project, linking them to a well-established measurement instrument which has proven to be

[1]For a general discussion, please see, e.g., Redman et al. (2021).

predictive of real-world behavioral impact. Second, the Fridays for Future (F4F) participation of the surveyed students was used to draw further conclusions about environmental activism resulting from a high environmental attitude. This in turn can be interpreted as an exemplary behavioral manifestation of pro-sustainable behavior for a more sustainable society, which consequently serves to achieve the ESD goals.

15.2 Theory

15.2.1 Defining Fitting Indicators for ESD in Schools: From Input to Outcome Orientation

In their review on the development of indicators for educational reporting, Baethge et al. (2010, p. 15) define indicators as quantitative tools (or proxy variables) providing a simple and comprehensible status report on the quality or the state of art of a more complex, usually multi-dimensional system. This is in line with the general direction of indicator definitions. However, in some fields, they are a lot more policy oriented as, for example, in Bormann and Michelsens' "the Collaborative Production of Meaningful Measure(ment)s" (2010), where indicators "are considered to provide condensed information which can be transformed into knowledge relevant for decision making." Also depending on the literature, different types of indicators are mentioned. In the framework of this analysis, we would like to apply the classification of Tilbury and Janousek (2006) and the Expert Groups on ESD Indicators as they are used by and for the ESD field, but equally offer connectivity to other fields, as this partly follows the standard international context/input-process-output-outcome scheme (see e.g., DIPF, 2007, p. 5):

- Checklist Indicators: provide information on initial policy, legislation, regulatory, and governance measures taken by a government in order to implement the Strategy.
- Input Indicators: provide information on a broader spectrum of activities taking place in terms of the implementation of the Strategy (e.g., amount of public authority money invested in the ESD materials, proportion of public supported research on ESD).
- Output Indicators: provide information on the direct results of these activities (e.g., performance of trained teachers, number of businesses involved in ESD projects, number of educators who received training on ESD issues).
- Outcome Indicators: provide information on the possible impact of the implementation of the Strategy particularly on values, attitudes, and choices in favor of SD (e.g., learning outcomes resulting from ESD partnerships, community-based projects, and business involvement). (Tilbury & Janousek, 2006, p. 11)

As stated above, before empirical measurement is possible, the competency dimensions of interest must be defined with sufficient precision. Such a specification allows, in principle, the operationalization of the competency of interest in an appropriate measurement procedure (see Klieme & Hartig, 2007). In general, there is a quite common acceptance of the broad tripartite classification of ESD goals in terms of achieving competencies in the knowledge, attitudes, and behavioral dimensions (e.g., Rieckmann, 2018; Waltner et al., 2019). The regional educational plan (implemented in 2016 for Baden-Wuerttemberg) served as a basis for the subject-specific ESD goal orientation. Teaching plans and curricula as well as school/ESD experts (i.e., teachers, students, university staff) were consulted in the identification of the different ESD goals for this specific field, the secondary schools. The ESD research program at Otto-von-Guericke University Magdeburg took another approach, where the ESD goal dimensions where defined based on a more international framework (the SDGs). These approaches are only two possibilities for defining the content to be measured.

No matter what the goal framework might be (regional, international – education plan, policy document) in any case of (outcome) measurement, precise goal (content) specifications and operationalization are needed in order to capture the underlying construct.

15.2.2 Operationalization of ESD Outcomes, Theoretical Background, and Disciplinary Connectivity

Without operationalization, the need for ESD and more importantly the outcomes of ESD-related interventions (e.g., lessons, seminars, projects) cannot be determined empirically (Gräsel et al., 2012; Wiek et al., 2011). Until now, sustainability and ESD research still lack connectivity to well-grounded models and findings from related relevant disciplines; these include environmental psychology, competence research, or more specific fields such as the science of behavior or attitude research (see, e.g., Waltner et al., 2019). Connecting ESD research to well-established measurement procedures facilitates the integration of already operationalized facets of competencies (e.g., environmental attitude) into the larger construct of SCs (sustainability competencies). The findings in measuring environmental attitudes are pertinent for ESD research which is aiming at measuring the broader dimensions of SCs – including, for example, the fields of sustainability-related knowledge and attitudes toward sustainability or sustainability-related problem-solving. SCs thus contain supplementary dimensions to the ecological dimension, such as political, economic, or more socially oriented considerations.

For the cognitive dimension of a SC, numerous attempts of operationalization in the form of knowledge scales are already available. Most of them precisely capture only specific parts of sustainability knowledge (e.g., environmental knowledge) as a significant subset of sustainability knowledge (e.g., Frick et al., 2004; Maloney & Ward, 1973; McBeth et al., 2011; Roczen et al., 2014). Thus, the cognitive dimension is not only about knowledge but also about cognitive skills and abilities (Waltner et al., 2019). However, although environmental knowledge is found to be consistently and positively related to environmental attitudes, the relationship is not especially strong (e.g., Arcury, 1990).

In another study, the behavior prediction through different forms of knowledge was also quite low: According to Frick et al. (2004), "the low overall explained behavioral variance of 6% was comparable to other studies (e.g., Hines et al., 1986/87). Although apparently small, this link should not be underestimated, since influences of knowledge on behavior are thought to be indirect, which means that they are mediated by other variables (cf. Kaiser & Fuhrer, 2003)." Thus, environmental knowledge can be regarded as the basis for ecological behavior of individuals, while a person's appreciation for nature is a relevant motivational factor for promoting the actual behavior (e.g., Kaiser et al., 2011).

Accordingly, for the affective-motivational and behavioral dimension, a very close connection has been proven by various studies. Affective goal commitment or a positively valued sequence of actions is the core of every motivation. An action is not carried out if the costs are perceived as too high when compared to the affective goal commitment. This attitude-cost relationship is modeled in the Campbell paradigm (Kaiser et al., 2010), which implies a solid link between a person's attitudes and his or her behavior. Consequently, in the framework of the Campbell paradigm, behavioral self-reports are used as indicators for a person's attitudes. Operationalization approaches of this attitude and behavior dimension can be found in various measurement instruments which have been broadly applied, for example, in Michelsen et al. (2015) (Greenpeace Sustainability Barometer), Shepherd et al. (2009) (Sustainable Development Values scale), Bogner (2018) (2-MEV scale), and Kaiser (1998) (GEB). According to the criteria of specific objectivity (Kaiser et al., 2018), in the future it might be preferable to focus on the subjacent level of the latent construct of the scales instead of fighting a battle of individual items. The basis for this claim will be elaborated in the section "Validation of New Indicators - Establishing Impact-Relevance of ESD Outcome."

15.3 Methods

In the following, the operationalization, the measurement process, and the data from the longitudinal assessment in secondary schools will be presented. A total of 1318 students and 113 teachers in grades 5–8 were surveyed at 10 randomly selected schools in the state of Baden-Württemberg (B-W). The proportion of school forms corresponds to the students' ratio at secondary schools in B-W and therefore constitutes a stratified sample according to this distribution of the Baden-Wuerttemberg Statistical Office (2017). All schools included in the study did not have a specific ESD program. Data was collected at the beginning and at the end of the school year 2018–2019. This longitudinal data can be used for answering the research question of the development of SC (i.e., the development within this 1 school year) and thus allows us to draw conclusions about the outcome of the ESD policies in B-W. For survey development, well-established items and findings from empirical educational, ESD, and environmental research as well as environmental psychology and sustainability psychology were combined with ESD-curricular and subject-specific items drawn from a curricular analysis for B-W. The survey development process is described in more detail in Waltner et al. (2019). On-site surveys were conducted with the assistance of teachers, who received verbal and written briefings prior to the survey, and four individuals from the research project. Participants were assured full confidentiality and anonymity by assignment of a code to match the data sets. The test duration was 90 min (with a break after approximately 45 min). Altogether, data was collected in 79 school classes (through whole class assessments). For more details concerning the pilot phases, the structure of the questionnaire, number of items, reliability, and validation, please see Waltner et al. (2019). The survey contained several SC dimensions such as sustainability knowledge, attitudes, and behavior. The three before-mentioned dimensions of SC will be presented in the following.

15.3.1 Sustainability Knowledge

Sustainability knowledge was measured with 16 multiple-choice items with one correct and three distractor answers to reduce guessing probability (compared to a true/false format). The questionnaire assessed knowledge about environmental protection, pollution, and other sustainability-related relevant topics, such as recycling, plastic in the oceans, erosion, and fair trade. The students' knowledge was operationalized by a mixture of basic, more general (cross-curricular) and subject-specific items. The general items were questions about common knowledge related to sustainable development. The subject-specific sustainability knowledge items were derived from the specific educational plans (see Ministry of Education Baden-Württemberg, 2016) for the different grades (5–8). A sample item is given below:

> We use a lot of plastic in our country. Why is this plastic bad for the fish in the ocean?
> Please mark only one answer (the most applicable one).
> - ☐ Because a lot of water is polluted during the production of plastic. The dirty water is led into the ocean and poisons the fish there.
> - ☐ Because large quantities of fish get caught in plastic bags, sticking the gills together and suffocating the fish.
> - ☐ The plastic often ends up in the habitat of the fish and rots there only very slowly. Sometimes the fish eat the plastic and can be injured.
> - ☐ Because the plastic factories pollute the air. When it rains, the dirt gets from the air into the water and damages the fish there.

15.3.2 Attitudes Towards Sustainability and Self-Reported Sustainability-Related Behavior

The two dimensions, affective-motivational beliefs toward sustainability and self-reported sustainability-related behavior, were measured using a four-point Likert-type scale: (1) strongly disagree; (2) disagree; (3) agree; and (4) strongly agree. An uneven scale was avoided, as indeci-

sive persons tend to choose a middle point (see, e.g., Chyung et al., 2017). These two scales consisted, respectively, of 16 evaluative verbal statements about the students' attitudes toward sustainability (e.g., *When I hear of cars that consume a lot of fuel, I get angry* or *the extinction of many animal and plant species makes me sad*) and of 13 self-reports about sustainability-related actions which the students could already perform or mostly decide themselves (e.g., *When I buy chocolate with my pocket money, I buy organic or fair-trade chocolate* or *On excursions, I take drinks in plastic bottles or disposable packaging,* e.g., *cans or PET bottles*).

15.4 Key Findings

In the framework of this chapter, we will not be able to describe in detail all findings which emerged from the project. We will, for example, not go into detail about the teacher-level data. However, the results on the development of SC at the student level are of course equally relevant for teachers, as they (can) influence the development of student competencies through appropriate instruction or educational actions. For the findings on the teacher level, see Waltner et al. (2020).

15.4.1 Sustainability Knowledge

Sustainability-related knowledge generally increases with age (Leeming et al., 1995; Waltner et al., 2019). We found this (positive) effect not only with increasing age across grade levels but also as an effect within the school year for each grade level separately. The sustainability knowledge increased statistically significantly across all grade levels within the school year ($F(1,1308) = 197.13$, $p < 0.001$); partial $\eta^2 = 0.13$; see Fig. 15.1.

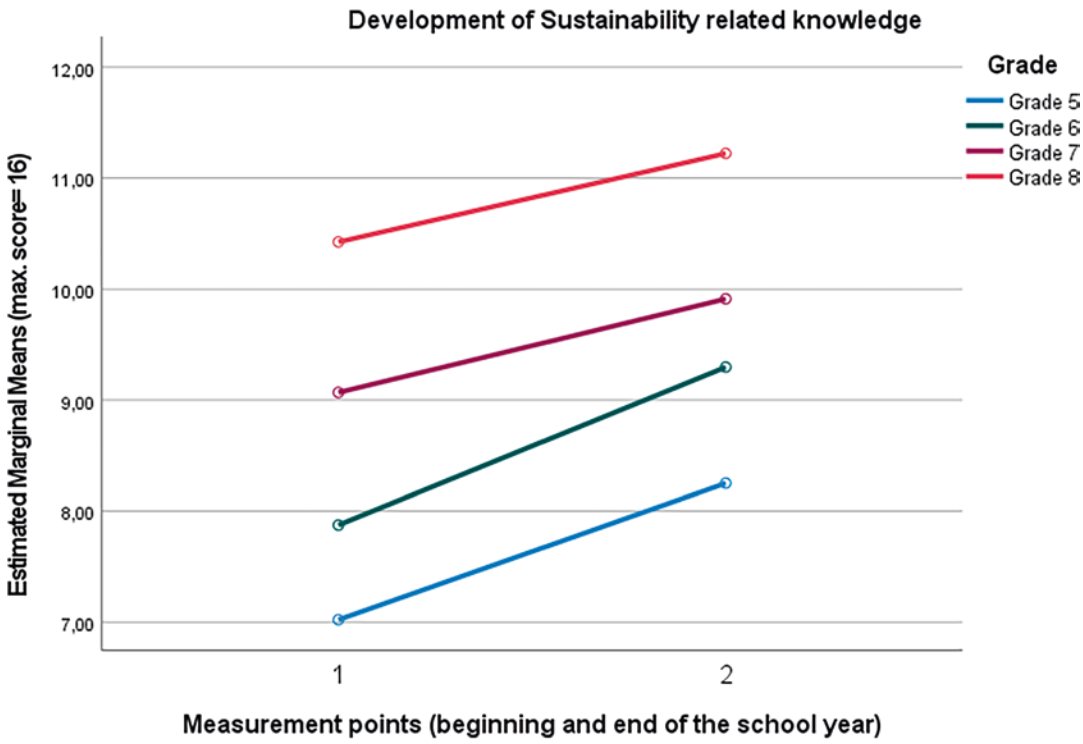

Fig. 15.1 Development of sustainability-related knowledge within 1 school year

15.4.2 Sustainability Attitudes

As mentioned above, there is a strong connection between sustainability attitudes and environmental attitudes which are a part of the broader concept of sustainability. Hence, the general tendencies should be comparable to findings of previous research in the field of environmental attitudes. According to previous research, younger children also tend to have a higher environmental attitude than older children (Krettenauer, 2017; Leeming et al., 1995; Liefländer et al., 2013). This holds true in our sample as well (see Fig. 15.2; grade 5 = lowest age group, grade 8 = highest age group). However, for the decreasing attitudes within 1 school year, other research findings are still scarce. Sustainability attitudes in our sample decreased statistically significantly across all grade levels within the school year ($F(1,1308) = 20{,}44$, $p < 0.001$); partial $\eta^2 = 0.02$; see Fig. 15.2.

15.4.3 Sustainability Behavior

Self-reported behavior showed either a very slight but not significant decrease or increase depending on the different grades (see Fig. 15.3). Changes are not statistically significant across all grades ($F(1,1318) = 0.21$, $p = 0.649$); partial $\eta^2 = 0.00$.

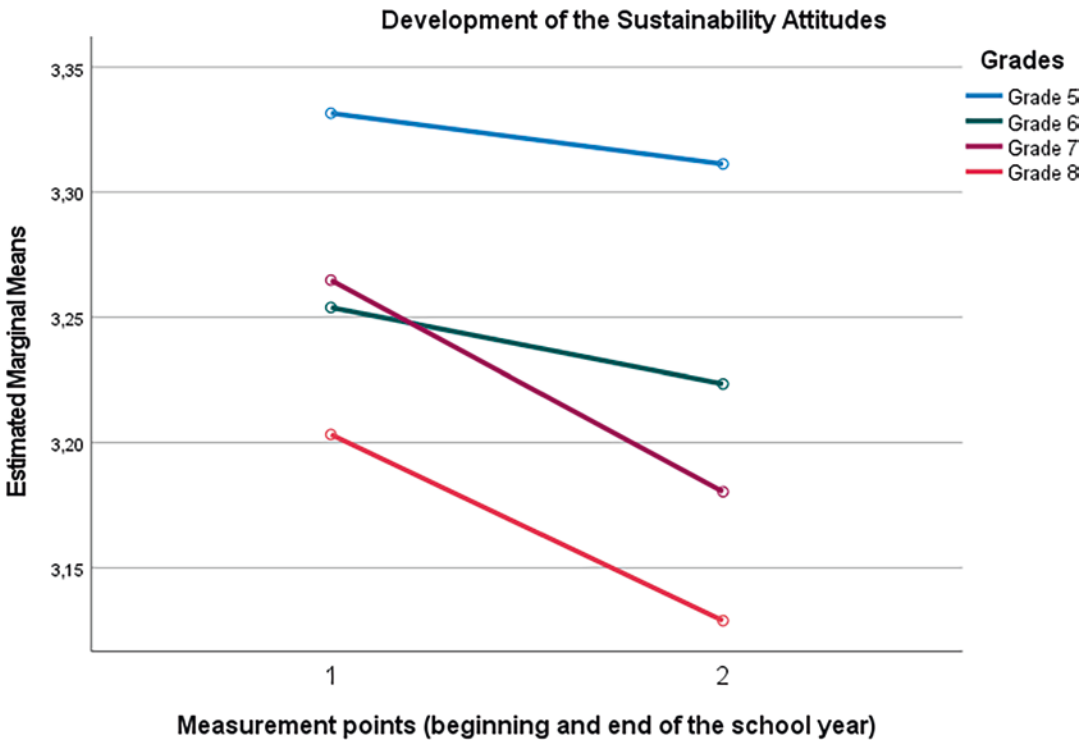

Fig. 15.2 Development of sustainability attitudes within 1 school year

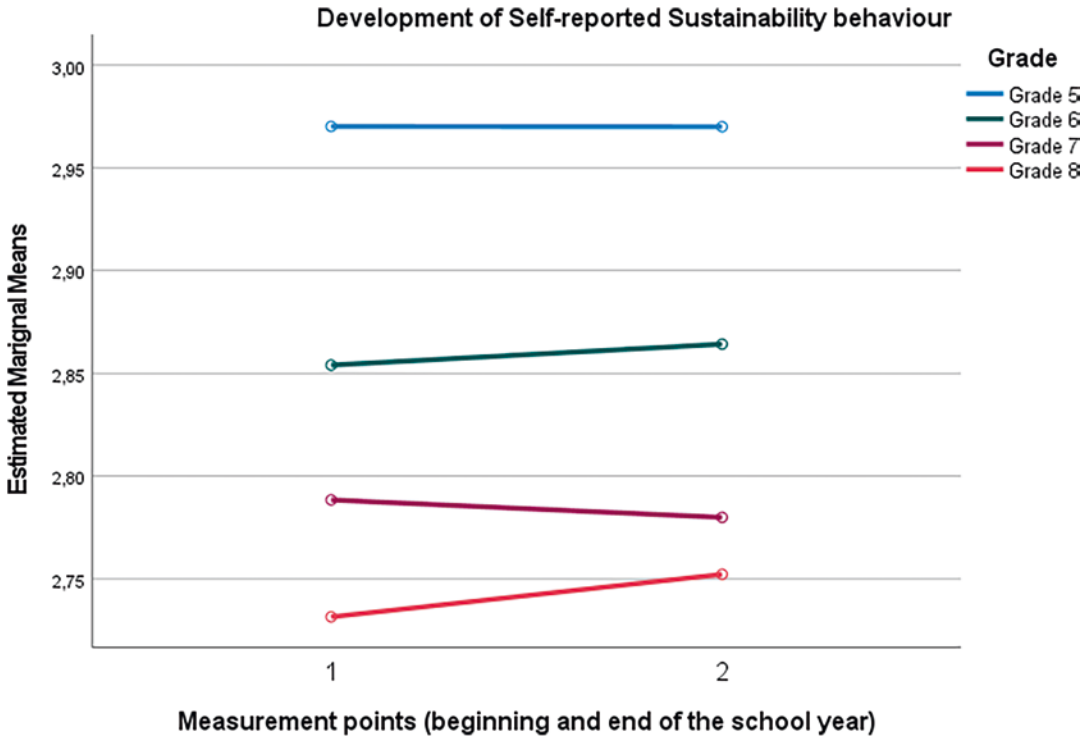

Fig. 15.3 Development of self-reported sustainability behavior within 1 school year

15.5 Validation of New Indicators-Establishing Impact-Relevance of ESD Outcome

> If governments and other social agencies are to heed what psychologists have to say on [the issue of environmental behavior], then the validity of our statements becomes very important. (Lloyd, 1994, p. 132)

If we wanted to make failure-proof assumptions about the impact of any intervention (such as ESD), we would have to consult observational data on concrete behavior. For example, say we observe that a student who used to leave the light on when leaving his or her room for dinner changes this behavior after receiving information about the environmental harm of the waste of power; in this case, we can assume that the gain in knowledge had an impact on the student's actual behavior and thus his or her impact on the environment. In reality, self-reports are the more frequently used indicators of behavior. Working with self-reports has two relevant advantages over observational data: first, they are easier to obtain, especially in large quantities; second (and based on the first advantage), they allow for a broader assessment of different behaviors (e.g., energy conservation behaviors, regional consumption behaviors, recycling behavior, etc.) at the same time and can be used to form aggregated measurements for behavioral classes (Kaiser et al., 2001). Research has shown, however, that there is a disparity between indicators based on self-reported behavior and objective (observable) behavior (e.g., Corral-Verdugo, 1997). The reasons for this have been widely discussed – a comprehensive review of this would go beyond the scope of this chapter. The main message which we want to draw from this insight is that indicators which are designed to assess the outcome of ESD have to be validated as to their congruence with real-life outcomes in the context of a shift to a more sustainable society or decisions; otherwise we might make wrong conclusions based on the data which we gather. This holds true for behavioral self-reports as well as for attitudinal

indicators and knowledge-based scales. Two exemplary efforts at validating the indicators used in the presented research project are shown in the following section. The underlying debate about whether or not the actual change of behavior can and should be a goal of ESD interventions will be referred to in the discussion section of this chapter.

15.5.1 Validation Through the Connection to Established Indicators

In a first attempt at validation, we performed a conceptual examination, using items from the newly developed ESD measurement project presented here and linking them to a well-established measurement instrument which has proven to be predictive of real-world behavioral impact: the General Ecological Behavior Scale (GEB) (see, e.g., Kaiser & Wilson, 2004). To do this, we assembled an 18-item ad hoc scale from the large survey, which was conducted within the survey presented above (see the Methods section above). Eight of these items were evaluative verbal statements (classic attitude items), and ten items were behavioral self-reports. Following the Campbell paradigm (Kaiser et al., 2010), both item types can be used in order to assess an underlying (latent) attitude, which motivates verbal as well as behavioral responses toward the attitude object. Within the Campbell paradigm, a person's attitude becomes transparent in the amount of behavioral cost said person is willing to overcome in order to pursue their goal (Byrka et al., 2017). So far, research on people's attitudes toward the environment has demonstrated that the Campbell paradigm – and thus its conceptual account of individual behavior – holds true for approximately 95% of the people in a given society (Kaiser et al., 2014). For more information about the derivation of attitudes from various types of manifest indicators attitudes (e.g., evaluative verbal statements and behavioral self-reports), see Kaiser and Wilson (2019).

The GEB uses 50 behavioral self-report items to assess attitudes toward the environment (Kaiser & Wilson, 2004). This operationalization has been used in countless studies (e.g., Evans et al., 2018; Overbeck & Kibbe, 2020; Otto & Kaiser, 2014), it has been translated into many languages, and it has proven to be predictive of real-life outcomes, such as the acceptance of nature preservation-related restrictions (Byrka et al., 2017) or a person's membership in an environmental organization (Arnold & Kaiser, 2018). We selected those (18) items for this validation attempt which we assumed would capture students' environmental attitudes as a substantial part of the attitudes toward sustainability concept as was fully assessed in the original questionnaire.

The goal of this validation was to show that the items which were newly developed within the project presented in this article prove to be valid predictors of environmental attitudes (as a subset of sustainability attitudes) – and in this way deduce that the development process of the project was successful. As a consequence, the competency differences which were assessed with this measurement instrument could point toward meaningful differences between the students which may have an actual impact on their future behavior. To investigate this, we conducted a validation study with $N = 154$ young adults (mean age = 30.21) in which we assessed the 18 items which we selected for our ad hoc measure alongside 47 GEB items. The correlation between the two measurements was $r = 0.75$ and when corrected for measurement error $r > 1$; this shows that the two scales measure the same latent construct, namely, environmental attitude. Additionally, we performed a joint Rasch calibration of all 65 items, which showed a person score reliability of rel. = 0.86 and a generally reasonable fit to the one-dimensional Rasch model; this indicates that all items assess the very same latent attribute, thus supporting our hypothesis.

On the other hand, our findings also provide support for the Campbell paradigm (see Kaiser et al., 2010) – in this paradigm, personal attitudes can be derived from verbal acts, such as expressions of appreciation for the environment and self-reports of past engagement in

environmentally friendly behaviors (Kaiser et al., 2018). Our findings also show that it is not relevant with which specific items a latent attitude is assessed but that any number of reasonably well-phrased behavioral or verbal self-reports which are aimed at the attitude object in question can be used to infer the underlying disposition (i.e., attitude). This supports the call for a higher priority of specific objectivity within the validation criteria for measurements in general (for a detailed account, see Kaiser et al., 2018).

15.5.2 Validation Through Prediction of Impact-Relevant Behavior

In our second step of validation, the (voluntary and self-determined) Fridays for Future (F4F) participation of the surveyed students is used to draw conclusions about environmental activism resulting from a high level of environmental attitudes. The participation in this activist movement can be interpreted as a behavioral manifestation aimed at promoting a more sustainable society, which consequently serves to achieve the ESD goals. The participation in the F4F movement was recorded at the second assessment point of the project (in the summer of 2019) with a simple question regarding this activity within the last year: "Have you heard of the Fridays for Future Movement" with the answering options "no"(N = 283), "yes" (N = 868), and "yes, and I have participated" (N = 163). In this case, the self-reported behavior was regarded as a valid proxy for actual behavior since self-reported behaviors seem to be especially reliable and valid when distinct facts are reported in a simple (dichotomous) way (e.g., Kaiser et al., 2003; Kormos & Gifford, 2014).

The data showed that students with higher sustainability attitudes in the first measurement had a significantly higher likelihood to participate in the F4F movement than students with a lower sustainability attitude. The same results could be shown for students' sustainability knowledge. Higher competency levels in both outcome variables could successfully predict the participation in a real-life, impact-relevant behavior, i.e., F4F. This validation through a criterion outside of the measurement process is a crucial step in gaining an understanding of any attribute one wants to measure (Whitely, 1977). We can be relatively confident that our measurements meet their goals; this is because the newly developed scales from the survey presented here can predict actual behavior which aims at the targets which the assessed competencies are supposed to support.

15.6 Limitations and Conclusions

> Not everything that can be counted counts and not everything that counts can be counted.
> Attributed to Albert Einstein. (see e.g., McKee, 2004)

Even if one can agree with this quotation in principle, it should be noted that non-measurable target and implementation criteria are often subject to the risk of remaining on the level of general normative statements or vague or unrealized implementation stages and policies without achieving the wanted outcome. The exemplary outcome indicators shown in this study, with the longitudinal data at the level of the pupils, provide ESD stakeholders with a useful information base (cf. e.g., DIPF, 2007; Oekes, 1989). Methodologically quantitative research projects can make a very important contribution to the normative debate, through empirical insights. Moreover, this debate also involves the normative questions of whether all dimensions have the same importance or weighting or whether some sub-dimensions should be weighted more heavily in a latent construct, depending on how they are assessed for scholastic or societal relevance. Recalling the educational policy definition of ESD given at the beginning of this paper, education for sustainable development enables learners to make informed decisions and act responsibly for the protection of the environment and for a functioning economy and a just world society for current and future generations (Ministry of Education Baden-Württemberg, 2016). In this

context, the acting component is clearly accentuated. So, one might ask the critical question: Do we need not only a shift from the input to the outcome orientation in the analytical/evaluative perspective but also a shift of attention from the purely cognitive to the behavioral components of SCs? Also, in terms of our research foci, we should not shy away from a call for more impact focused research. Indeed, the call for more impact focus within the sustainability sciences, such as environmental psychology, is getting more prominent (see, e.g., Nielsen et al., 2021).

15.6.1 Why Is Measuring Relevant for Teachers?

Studies on ESD implementation processes might help to further illuminate the interconnections between seemingly different realities of policy-makers and "the street ministers of education" (i.e., the teachers). In terms of Lipsky and other bottom-up theorists, they are also policy-makers "on the ground" as they are the very important stakeholders who implement the policies, i.e., in the classrooms (Lipsky, 2010; Waltner et al., 2020). In this way, the actors (in the sense of the bottom-up approach, also including teachers and students themselves or any ESD learners) can initiate desired developments in a more empirically based and targeted manner and do not only have to refer to plausible and normative considerations and assumptions. Decisions can therefore be taken based on empirical long-term data and do not only need to be based on intuitively based perceptions or observations which the teachers make in the classroom. Furthermore, this paper reported the development and validation of a measurement tool for an interdisciplinary SC. It is obvious that, in the further course of research, the specific contributions of the different subjects (biology, geography, economics, politics, social studies, etc.) to ESD should also be considered. Accordingly, instruments for facets of a subject-specific SC would then have to be developed and tested (Rieß et al., 2018; Waltner et al., 2019).

Our research adds to a growing understanding of SCs, their development, and the sustainability and educational governance through policy-making. On this basis, appropriate evidence-based recommendations for the further development of ESD research and the implementation of ESD in school practice can be formulated. Through the possibilities of measurement presented and the data already generated, further insights into the successful implementation of ESD in schools and the associated conditions for success can be gained. Despite these advantages, some shortcomings should be mentioned. First, as implicated by the general definition of indicators, they only offer an approximation to the real world. Even though, as described in the section "Validation of New Indicators-Establishing Impact-Relevance of ESD outcome," we tried to validate our ESD outcome measures, they still remain a mere proxy for reality.

A second point which is important to mention in considerations about the limitations of indicators – and especially the wish to globally assess and compare the results of ESD or SDG outcomes – is the aspect of a reduced context sensitivity. Even for our study only focusing on one state in Germany, we needed to deal with the trade-off between regional specificity (or the aim of a detailed analysis of all the possible contexts and possible factors influencing SC) and the aim of getting a broad picture of the state of play and development. With the aim of a general assessment tool of ESD outcomes, e.g., in a global scope, this trade-off would become even more evident as the possibilities of observing all the details shrink. When aiming for global indicators which are easily replicable and comparable, detailed observations need to be sacrificed. Consequently, additional qualitative studies looking into further details of the SC development process would represent a useful complement.

We can currently still state, in agreement with Michelsen et al. (2012), that interlinkages hardly exist even in the fields of youth, learning, and education which are closer to ESD research. This holds true, not only within the different methodological fields but also in the different research domains, with regard to the networking and cooperation between different research and educational areas. Based on the experience of this

collaboration, we would recommend a stronger interlinking of the already mentioned relevant scientific disciplines (such as (educational) psychology, educational science, environmental and sustainability sciences). Furthermore, a closer integration of experts from practice (pupils, teachers, students, etc.) with regard to the findings of ESD research which are important for school education – but also for society as a whole – seems to be an important goal.

In view of the impending global challenges, however, a global society which is serious about achieving the SDGs cannot afford to remain at the level of normative statements and vague target formulations. The corresponding possibilities of measurability will ideally stimulate benchmarking processes (concrete target formulations) (e.g., Sayed & Asmuss, 2013). An international monitoring system for the ESD implementation efforts of the various countries would be purposeful, in view of the future global challenges for humankind (Independent Group of Scientists appointed by the Secretary-General, 2019). Such an international orientation would facilitate the exchange of insights in the implementation processes of ESD and would also offer possibilities for international cooperation (see, e.g., Reid et al., 2006).

Acknowledgments The authors gratefully acknowledge all project members of the research groups and other supporters. In addition, we would like to thank the school principals, contact persons, and all other school stakeholders for their valuable contributions during the different project phases.

Funding This research was partly funded by the Ministry of Education, Youth, and Sports of Baden-Wuerttemberg, the Ministry of Environment of Baden-Wuerttemberg, and the Stiftung Naturschutzfonds [*Nature Conservation Foundation*] (2-8802.00-BNE/50 and 73-8831.21/54691-1743L).

References

Arcury, T. (1990). Environmental attitude and environmental knowledge. *Human Organization, 49*(4), 300–304. https://doi.org/10.17730/humo.49.4.y6135676n433r880

Arnold, O., & Kaiser, F. G. (2018). Understanding the foot-in-the-door effect as a pseudo-effect from the perspective of the Campbell paradigm. *International Journal of Psychology, 53*(2), 157–165. https://doi.org/10.1002/ijop.12289

Baethge, M., Brunke, J., Döbert, H., Fest, M., Freitag, H.-W., Fitzsch, B., Fuchs-Rechlin, K., Kerst, C., Kühne, S., Scharfe, S., Skripski, B., Wieck, M., & Wolter, A. (2010). Indikatorenentwicklung für den nationalen Bildungsbericht "Bildung in Deutschland": *Grundlagen, Ergebnisse, Perspektiven* [*Development of indicators for the national education report "Education in Germany": Foundations, results, perspectives*]. BMBF.

Bogner, F. X. (2018). Environmental values (2-MEV) and appreciation of nature. *Sustainability, 10*(2), 1–10. https://doi.org/10.3390/su10020350

Bormann, I., & Michelsen, G. (2010). The collaborative production of meaningful measure(ment)s: Preliminary insights into a work in progress. *European Educational Research Journal, 9*(4), 510–518.

Burford, G., Tamás, P., & Harder, M. (2016). Can we improve indicator design for complex sustainable development goals? A comparison of a values-based and conventional approach. *Sustainability, 8*(9), 1–38. https://doi.org/10.3390/su8090861

Byrka, K., Kaiser, F. G., & Olko, J. (2017). Understanding the acceptance of nature-preservation-related restrictions as the result of the compensatory effects of environmental attitude and behavioral costs. *Environment and Behavior, 49*(5), 487–508. https://doi.org/10.1177/0013916516653638

Cai, Y., Lenton, T. M., & Lontzek, T. S. (2016). Risk of multiple interacting tipping points should encourage rapid CO_2 emission reduction. *Nature Climate Change, 6*(5), 520–525. https://doi.org/10.1038/nclimate2964

Chyung, S. Y. Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the likert scale. *Performance Improvement, 56*(10), 15–23. https://doi.org/10.1002/pfi.21727

Corral-Verdugo, V. (1997). Dual 'realities' of conservation behavior: Self-reports vs observations of re-use and recycling behavior. *Journal of Environmental Psychology, 17*, 135–145.

DIPF- Leibniz-Institut für Bildungsforschung und Bildungsinformation. (2007). Das weiterentwickelte Indikatorenkonzept der Bildungsberichterstattung [*The advanced indicator concept of educational reporting*]. http://www.bildungsbericht.de/de/forschungsdesign/pdf-grundlagen/indikatorenkonzept.pdf. Accessed 10 June 2021.

Evans, G. W., Otto, S., & Kaiser, F. G. (2018). Childhood origins of young adult environmental behavior. *Psychological Science, 29*(5), 679–687. https://doi.org/10.1177/0956797617741894

Frick, J., Kaiser, F. G., & Wilson, M. (2004). Environmental knowledge and conservation behavior: Exploring prevalence and structure in a representative sample.

Personality and Individual Differences, 37(8), 1597–1613. https://doi.org/10.1016/j.paid.2004.02.015

Gräsel, C., Bormann, I., Schütte, K., Trempler, K., Fischbach, R., & Asseburg, R. (2012). Perspektiven der Forschung im Bereich Bildung für nachhaltige Entwicklung [*Prospects for research in education for sustainable development*]. In Bundesministerium für Bildung und Forschung (Ed.), *Bildung: Vol. 39. Bildung für nachhaltige Entwicklung: Beiträge der Bildungsforschung* (pp. 7–25). Bundesministerium für Bildung und Forschung (BMBF) Referat Bildungsforschung.

Hines, J. M., Hungerford, H. R., & Tomera, A. N. (1986/87). Analysis and synthesis of research on responsible environmental behavior: A meta-analysis. *Journal of Environmental Education, 18*(2), 1–8.

Independent Group of Scientists appointed by the Secretary-General. (2019). The future is now. Science for achieving sustainable development. Global sustainable development report. https://sustainabledevelopment.un.org/content/documents/24797GSDR_report_2019.pdf. Accessed 14 June 2021.

Kaiser, F. G. (1998). A general measure of ecological behavior. *Journal of Applied Social Psychology, 28*(5), 395–422. https://doi.org/10.1111/j.1559-1816.1998.tb01712.x

Kaiser, F. G., & Fuhrer, U. (2003). Ecological behavior's dependency on different forms of knowledge. *Journal of Environmental Psychology, 52*(4), 598–613. https://doi.org/10.1111/1464-0597.00153

Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences, 36*(7), 1531–1544. https://doi.org/10.1016/j.paid.2003.06.003

Kaiser, F. G., & Wilson, M. (2019). The Campbell Paradigm as a behavior-predictive reinterpretation of the classical tripartite model of attitudes. *European Psychologist, 24*(4), 359–374. https://doi.org/10.1027/1016-9040/a000364

Kaiser, F. G., Frick, J., & Stoll-Kleemann, S. (2001). Zur Angemessenheit selbstberichteten Verhaltens: Eine Validitätsuntersuchung der Skala Allgemeinen Ökologischen Verhaltens [*On the appropriateness of self-reported behavior: A validity investigation of the General Ecological Behavior scale*]. *Diagnostica, 47*(2), 88–95. https://doi.org/10.1026//0012-1924.47.2.88

Kaiser, F. G., Doka, G., Hofstetter, P., & Ranney, M. A. (2003). Ecological behavior and its environmental consequences: A life cycle assessment of a self-report measure. *Journal of Environmental Psychology, 23*(1), 11–20. https://doi.org/10.1016/S0272-4944(02)00075-0

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review, 14*, 351–367. https://doi.org/10.1177/1088868310366452

Kaiser, F. G., Hartig, T., Brügger, A., & Duvier, C. (2011). Environmental protection and nature as distinct attitudinal objects. *Environment and Behavior, 45*(3), 369–398. https://doi.org/10.1177/0013916511422444

Kaiser, F. G., Arnold, O., & Otto, S. (2014). Attitudes and defaults save lives and protect the environment jointly and compensatorily: Understanding the behavioral efficacy of nudges and other structural interventions. *Behavioral Sciences (Basel, Switzerland), 4*, 202–212. https://doi.org/10.3390/bs4030202

Kaiser, F. G., Merten, M., & Wetzel, E. (2018). How do we know we are measuring environmental attitude? Specific objectivity as the formal validation criterion for measures of latent attributes. *Journal of Environmental Psychology, 55*, 139–146. https://doi.org/10.1016/j.jenvp.2018.01.003

Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften im erziehungswissenschaftlichen Diskurs [*Competencies' concepts in social sciences in educational discourse*]. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Zeitschrift für Erziehungswissenschaft Sonderheft: Vol. 8. Kompetenzdiagnostik* (pp. 11–31). VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden.

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology, 40*, 359–371. https://doi.org/10.1016/j.jenvp.2014.09.003

Krettenauer, T. (2017). Pro-environmental behavior and adolescent moral development. *Journal of Research on Adolescence, 27*(3), 581–593. https://doi.org/10.1111/jora.12300

Leeming, F. C., Dwyer, W. O., & Bracken, B. A. (1995). Children's environmental attitude and knowledge scale: Construction and validation. *The Journal of Environmental Education, 26*(3), 22–31. https://doi.org/10.1080/00958964.1995.9941442

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the earth's climate system. *Proceedings of the National Academy of Sciences, 105*(6), 1786–1793. https://doi.org/10.1073/pnas.0705414105

Liefländer, A. K., Fröhlich, G., Bogner, F. X., & Schultz, P. W. (2013). Promoting connectedness with nature through environmental education. *Environmental Education Research, 19*(3), 370–384. https://doi.org/10.1080/13504622.2012.697545

Lipsky, M. (2010). *Street-level Bureaucracy: 30th Ann. Ed.: Dilemmas of the individual in public service* (30th Ann. Ed). Russell Sage Foundation.

Lloyd, K. E. (1994). Do as I say, not as I do. *The Behavior Analyst, 17*(1), 131–139.

Maloney, M. P., & Ward, M. P. (1973). Ecology: Let's hear from the people: An objective scale for the measurement of ecological attitudes and knowledge. *American Psychologist, 28*, 583–586.

McBeth, W., Hungerford, H. R., Marcinkowski, T., Volk, T. L., & Cifranick, K. (2011). *National environmental literacy assessment, phase two: Measuring the effectiveness of North American environmental education programs with respect to the parameters of environmental literacy: Final research report.* North American Association for Environmental Education. https://www.noaa.gov/sites/default/files/atoms/files/NELA_Phase_Two_Report_020711.pdf. Accessed 10 June 2021.

McKee, M. (2004). Not everything that counts can be counted; not everything that can be counted counts. *BMJ (Clinical Research Ed.), 328*(7432), 153. https://doi.org/10.1136/bmj.328.7432.153

Michelsen, G., Grunenberg, H., & Mader, C. (2012). Greenpeace Nachhaltigkeitsbarometer: Was bewegt die Jugend? [*Greenpeace sustainability barometer: What affects young people?*]. VAS-Verlag.

Michelsen, G., Grunenberg, H., Mader, C., & Barth, M. (2015). Greenpeace Nachhaltigkeitsbarometer 2015: Nachhaltigkeit bewegt die jüngere Generation [*Greenpeace sustainability barometer 2015: Sustainability affects the younger generation*]. VAS-Verlag.

Ministerium für Kultus, Jugend und Sport Baden-Württemberg. (2016). Bildungspläne Baden-Württemberg: Bildung für nachhaltige Entwicklung (BNE). [*Education plans for Baden-Württemberg: Education for Sustainable Development (ESD)*]. http://www.bildungsplaene-bw.de/Lde/Startseite/BP2016BW_ALLG/BP2016BW_ALLG_LP_BNE. Accessed 10 June 2021.

Nielsen, K. S., Cologna, V., Lange, F., Brick, C., & Stern, P. C. (2021). The case for impact-focused environmental psychology. *Journal of Environmental Psychology.* Advance online publication. https://doi.org/10.31234/osf.io/w39c5

Oekes, J. (1989). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis, 11*(2), 181–199.

Otto, S., & Kaiser, F. G. (2014). Ecological behavior across the lifespan: Why environmentalism increases as people grow older. *Journal of Environmental Psychology, 40*, 331–338. https://doi.org/10.1016/j.jenvp.2014.08.004

Overbeck, A. K., & Kibbe, A. (2020). Decoding activism: Examining the influence of environmental attitude and proactivity on environmental activism. *Umweltpsychologie, 24*(1), 183–190.

Redman, A., Wiek, A., & Barth, M. (2021). Current practice of assessing students' sustainability competencies: A review of tools. *Sustainability Science, 16*(1), 117–135. https://doi.org/10.1007/s11625-020-00855-1

Reid, A., Nikel, J., & Scott, W. (2006). *Indicators for Education for Sustainable Development: A report on perspectives, challenges and progress*. https://researchportal.bath.ac.uk/en/publications/indicators-for-education-for-sustainable-development-a-report-on. Accessed 13 June 2021.

Rieckmann, M. (2018). Chapter 2. Learning to transform the world: Key competencies in ESD. In UNESCO (Ed.), *UNESCO's education on the move series. Issues and trends in education for sustainable development* (pp. 39–59). UNESCO.

Rieß, W., Mischo, C., & Waltner, E.-M. (2018). Ziele einer Bildung für nachhaltige Entwicklung in Schule und Hochschule: Auf dem Weg zu empirisch überprüfbaren Kompetenzen [*The goals of education for sustainable development in schools and universities: Towards empirically verifiable competencies*]. *GAIA - Ecological Perspectives for Science and Society, 27*(3), 298–305. https://doi.org/10.14512/gaia.27.3.10.

Roczen, N., Kaiser, F. G., Bogner, F. X., & Wilson, M. (2014). A competence model for environmental education. *Environment and Behavior, 46*(8), 972–992. https://doi.org/10.1177/0013916513492416

Sayed, A., & Asmuss, M. (2013). Benchmarking tools for assessing and tracking sustainability in higher educational institutions. *International Journal of Sustainability in Higher Education, 14*(4), 449–465. https://doi.org/10.1108/IJSHE-08-2011-0052

Schellnhuber, H. J. (2006). Facing climate change: Tipping points and u-turns. In D. Buckland (Ed.), *Burning ice: Art & climate change; the Cape Farewell project was created by the artist* (pp. 112–113). Gaia Project.

Shepherd, D. A., Kuskova, V., & Patzelt, H. (2009). Measuring the values that underlie sustainable development: The development of a valid scale. *Journal of Economic Psychology, 30*(2), 246–256.

Statistisches Landesamt Baden-Württemberg. (2017). Allgemeinbildende Schulen in Baden-Württemberg [*General education schools in Baden-Württemberg*]. http://www.statistik.baden-wuerttemberg.de/Service/Veroeff/Statistik_AKTUELL/803417004.pdf . Accessed 10 June 2021.

Tilbury, D., & Janousek, S. (2006). *Development of a national approach to monitoring, assessment and reporting on the decade of education for sustainable development: Stage 1: Identification of national indicators* [Summarising documented experiences on the development of ESD indicators and networking with expert groups on ESD indicators]. Australian Research Institute of Education for Sustainability and Australian Government Department of the Environment and Water Resources. http://aries.mq.edu.au/projects/esdIndicators/files/ESDIndicators_Feb07.pdf. Accessed 14 June 2021.

UNESCO. (2017). *Education for sustainable development goals: Learning objectives*. https://www.unesco.de/sites/default/files/2018-08/unesco_education_for_sustainable_development_goals.pdf. Accessed 14 June 2021.

Waltner, E.-M., Rieß, W., & Brock, A. (2018). Development of an ESD indicator for teacher training and the national monitoring for ESD implementation in Germany. *Sustainability, 10*(7), 2508. https://doi.org/10.3390/su10072508

Waltner, E.-M., Rieß, W., & Mischo, C. (2019). Development and validation of an instrument for measuring student sustainability competencies. *Sustainability, 11*(6), 1717. https://doi.org/10.3390/su11061717

Waltner, E.-M., Scharenberg, K., Hörsch, C., & Rieß, W. (2020). What teachers think and know about education for sustainable development and how they implement it in class. *Sustainability, 12*(4), 1690. https://doi.org/10.3390/su12041690

Whitely, S. E. (1977). Models, meanings and misunderstandings: Some issues in applying Rasch's theory. *Journal of Educational Measurement, 14*(3), 227–235.

Wiek, A., Withycombe, L., & Redman, C. L. (2011). Key competencies in sustainability: A reference framework for academic program development. *Sustainability Science, 6*(2), 203–218.

Eva-Maria Waltner studied at the University of Constance (Germany), Université du Québec à Montréal (Canada), and Université de Liège (Belgium), and graduated in political science, French, and pedagogy. She worked as an assistant for UNEP in Brussels and also in practical outdoor education. Eva-Maria completed her PhD in the project The implementation of Education for Sustainable Development (ESD) – a multilevel analysis of Sustainability Competencies (SC) in schools at the University of Education Freiburg (Germany) and is currently involved in the international The Monitoring and Evaluating Climate Communication and Education Project (MECCE).

Anne Overbeck is a member of the Division of Personality and Social Psychology at Otto-von-Guericke University Magdeburg. She studied psychology at Mannheim University (BSc) and later continued her studies at Magdeburg University (MSc, focus on environmental psychology). Her main scientific interests are sustainability, personality-psychology, and political participation. Anne has been working on research concerning the outcome-oriented assessment of ESD in students. In her PhD studies, she concentrates her research on the determinants of environmental activism.

Werner Rieß is Professor of Biological Education at the University of Education Freiburg (Germany). In his research projects, he particularly investigates the effects of ESD from kindergarten to university on systems thinking and other important goals (e.g., knowledge, attitudes, competencies, behavior of learners of different ages). In addition, his research examines the current state of ESD with the aim of creating indicators for ESD.