

Review

Assessing the Impact of ESD: Methods, Challenges, Results

Quentin Sossé ^{1,*} , Johanna Wagner ^{2,*} and Carina Hopper ^{3,*}¹ Independent Researcher, 75020 Paris, France² Independent Researcher, 75015 Paris, France³ Independent Researcher, 08028 Barcelona, Spain

* Correspondence: quentin.sosse@hec.edu (Q.S.); johanna.wagner@essec.edu (J.W.); carina.hopper@essec.edu (C.H.)

Abstract: Education for Sustainable Development (ESD; also often called Education for Sustainability (Efs)) is a key lever of the United Nations' Sustainable Development Goals, which emphasize the need for everyone to have the knowledge and skills to meet the challenges of creating a more sustainable world. However, while we can find examples of ESD across the globe, its potential to scale effectively and its impact on achieving the goals of sustainable development as compared with traditional curricula are often questioned. This literature review, at the crossroads of econometrics, educational sciences and psychology, aims to foster scaled ESD research and initiatives by offering a better understanding of the doubts that surround its potential impact. To that end, we (1) shed light on the methods and good practices for assessing this impact; (2) underline the specificity of the data to be collected in the context of these methods of assessment; and (3) outline the existing conclusions of impact studies dedicated to ESD that have served to highlight the limits and challenges for accurate measurement. These impact studies suggest that ESD will only achieve its objectives if pedagogical approaches are renewed. The inclusion of studies showing de facto poor results for ESD makes it possible to complete the picture of the endogenous and exogenous factors determining sustainable behavior that must be taken into account, both in the design of impact assessment tools and in the concrete implementation of ESD.

Keywords: education for sustainable development; sustainability education; environmental literacy; competences for sustainable development; sustainability citizens; impact assessment; higher education curricula



Citation: Sossé, Q.; Wagner, J.; Hopper, C. Assessing the Impact of ESD: Methods, Challenges, Results. *Sustainability* **2021**, *13*, 2854. <https://doi.org/10.3390/su13052854>

Academic Editors: Lela Mélon, Roberto Caranta and Carlos Ignacio Gómez Ligüerre

Received: 1 February 2021

Accepted: 4 March 2021

Published: 6 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. The Origin of ESD in the Wake of Early Environmental Concerns

Although ecological thinking predates this era, it was the 1970s that brought about a real “disinhibition” [1] of environmental consciousness that became widespread and global. In 1972 and after 30 years of unprecedented Western economic growth, the Club of Rome’s “The Limits to Growth” [2] report already seemed to sound the alarm about the exponential depletion of natural resources. However, it was not until 15 years later with the Brundtland Report [3] that the concept of “sustainable development” (SD) emerged as “development that aims to meet the needs of present generations without compromising the ability of future generations to meet their own needs”.

Political ecology was born, and with it the idea of international supervision of these issues. Capitalizing on its overarching role, it was the UN that first seized the concept and consolidated SD as a key element of its long-term agenda. At the Rio Earth Summit in 1992, when the concept was already widely popularized, the UN clarified its definition and added to the environmental imperative the criteria of social justice and economic progress, which together form the three pillars of SD.

Spearheading of SD by international bodies must not, however, allow us to lose sight of the civic responsibility that lies at the heart of these issues. For, if reports extensively

point to anthropogenic responsibility for the disruptions at work [4], it is because it is a responsibility shared by the more than seven billion people whose current lifestyles and behaviors are incompatible with a sustainable future.

There is no doubt that, despite efforts made at the international and national levels, SD can only be achieved through a profound change in the way people think and act. This change will have to be organized at the global level but most importantly at the individual level in order to equip everyone with the knowledge, values and skills of “sustainability citizens” [5]. The conclusion is clear: SD is a goal that can only be achieved through behavioral change.

At the heart of this imperative, education is seen as a crucial element in the advent of a more sustainable world today and tomorrow. Yet, as UNESCO reminds us [6], “not all kinds of education support sustainable development. Education that promotes economic growth alone may well also lead to an increase in unsustainable consumption patterns”. As noted by Orr [7], “the kind of education we need begins with the recognition that the crisis of global ecology is first and foremost a crisis of values, ideas, perspectives, and knowledge, which makes it a crisis of education, not one in education”. Behind these warnings lies a clear idea: the reform represented by SD can only be achieved through an equally reforming approach in education.

It was in the 1960s, in the wake of the institutionalization of environmental issues, that the corpus of Environmental Education (EE) was first born. The UNESCO Biosphere Conference of 1968, and later on the Belgrade Charter (1976)—issued after the Belgrade working Conference on Education (1975) and the Tbilisi Conference (1977)—set out the first theoretical framework of EE [8], defined as “the process aimed at developing a world population that is aware of and concerned about the total environment and its associated problems, and has the attitudes, motivations, knowledge, commitment and skills to work individually and collectively towards solutions of current problems and the prevention of new ones”.

As raised by Kopnina [9], the initial ambiguity behind the term “total environment” often confined EE to aspects of nature conservation. In 1992, however, Agenda 21, which emerged from the Rio Earth Summit negotiations, introduced a discourse of SD and its three pillars into the educational reform of EE, which would gradually be supplanted in practice by the corpus of “Education for Sustainable Development” (ESD).

1.2. Theoretical Foundations of ESD: Holism of Contents, Pluralism of Pedagogical Approaches

As a true pedagogical innovation aimed at empowering citizens to take action and make decisions that are committed, sustainable and responsible towards the Earth, people and economic systems, ESD (and EE) differs from traditional modes of education thanks to two interdependent theoretical axioms first highlighted by UNESCO that underpin its *raison d'être*, the first dealing with the content of ESD, which should be “holistic” [6] (“what” should be taught), and the second with the pedagogy, which should be “pluralistic” [6] (the “how”). These two facets are, moreover, well known in the literature: “ESD continues to grow both in content and pedagogy and its visibility and respect have grown in parallel” [10].

The theoretical content of ESD, first of all, addresses a variety of topics and disciplines. As specified in the UNESCO definition, “ESD calls on countries to ensure that all learners are provided with the knowledge and skills to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture’s contribution to sustainable development” [6]. ESD therefore aims at interdisciplinarity, a holistic approach defined by Ohman [11] as education that is able to integrate the multiple perspectives of the three pillars of SD, emphasizing their interactions and contingencies, in time and space, and at the local, regional and global level.

As much as its content, which must be “holistic”, the pedagogical approach of ESD constitutes its second distinctive element and will be “pluralistic”. The aim is to train students to recognize and integrate different perspectives, ways of being and values in order to equip them with the skills needed to take effective action to meet the challenges of today and tomorrow. Behind this axiom lies the very idea of the interdependence and complexity of the issues at stake, so great that it cannot be effective to try to introduce them independently and teach predefined solutions. Indeed, the pluralism of ESD calls for a focus on reflection around these issues rather than on the teaching of “right answers”. To that end, ESD will need to be accompanied by new pedagogies that encourage action in different environments and enable young citizens to understand the world through their own observation and to develop skills for sustainable awareness and behavior [12,13].

According to UNESCO, which is at the origin of these two predicates, holism and pluralism are generally seen as intrinsically linked, in the sense that learning about all aspects of SD can only take place in the pluralist understanding of the social construct, economic perspectives and environmental balance.

However, beyond these two axioms that structure curricula, it is the student experience as a whole that ESD calls to reform. Indeed, as Cortese [14] points out, “students learn from everything around them, these activities form a complex web of experience and learning”.

Thus, apart from pure learning activities, whether formal or non-formal, ESD also aims at reforming the whole range of activities that make up the ecosystem of the school, or university, and beyond. This will involve acting on learning modules, research, operations and communication at the same time, eventually extending ESD to spheres other than those of traditional education structures.

On the whole, a consensus is emerging: rather than a simple incremental innovation, ESD must therefore be part of an epistemological and pedagogical break with the past, both in its substance and in its form. It will be a question of overcoming the initial “epistemological error” [15], which considered man above and at the center of nature (anthropocentrism), and bringing about the ideal of a humanity thought to be within it (ecocentrism).

It was thus underpinned by these first conceptual principles that ESD was going to expand to spheres other than those of theory.

1.3. Acceleration and the First Roadblocks

As early as 2005, the UNESCO-initiated Decade of Education for Sustainable Development (DESD) launched a general movement in two phases:

- Until 2008 there would be a focus on defining and promoting ESD, in particular by mapping a network of relevant actors and partners;
- In 2009 and after the Bonn Declaration, ESD would be proactively implemented, notably in the UN Regional Centers of Expertise, which are pilot ecosystems on these issues.

In 2014, the Aichi-Nagoya Declaration that closed the Decade also opened the cycle of the Global Action Program, which extended the commitments to lead ESD by setting itself the mission of scaling up ESD actions and good practices tested during the Decade.

However, it is really only the following year that marks a clear turning point for ESD both conceptually and factually: in 2015 the concept became a strategic tool for transition. Indeed, in the wake of the Rio+20 Earth Summit in 2012, the 2030 Agenda for SD was adopted and placed at the forefront of its approach, with the 17 Sustainable Development Goals (SDGs) designed to overcome the barriers to SD (inequalities, consumption patterns, institutional fragility, environmental degradation) [6] in order to ensure a sustainable, prosperous and equitable future for all.

Placed at the core of the SDG framework, ESD has a leading role to play in a dual capacity:

- As a full-fledged SDG, behind Goal 4.7: “By 2030, ensure that all students acquire the knowledge and skills necessary to promote sustainable development, including through education for development and sustainable lifestyles, human rights, gender equality, the promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and the contribution of culture to sustainable development” [6];
- As an SDG at the service of others, accelerating progress towards achieving the overall SDG framework and contributing to the strategies that aim to achieve each of them. Thus, UNESCO stresses that “The SDGs, targets and means of implementation are thought of as universal, indivisible and interlinked. Each of the 17 goals has a set of targets. In each set, at least one target involves learning, training, educating or at the very least raising awareness of core sustainable development issues” [6].

From the laying of the first conceptual foundations of ESD to the closing of the DESD, concrete initiatives contributing to Goal 4.7 have flourished all around the world. Illustrating this trend, UNESCO’s Regional Centers of Expertise (RCEs), as pilot ecosystems for these new pedagogies globally, are leading the way and acting as relays for spontaneous initiatives at a more local/national level. Let us also mention the various initiatives labelling the schools spearheading ESD that have been set up as a network: “eco-schools”, the Association for the Advancement of Sustainability Education (AASHE), etc.

However, we are still far from the anticipated large-scale reform. For, apart from the RCEs, where it is given center stage, ESD is struggling to establish itself in its most complete form [16–18]: most of the time, it is barely relegated to a supplementary role in an education system that is still everywhere anchored in previous paradigms.

Lack of awareness among educational teams and stakeholders, the financial cost of the reform, lack of consensus on the methods and essence of ESD, etc., are all factors that prevent it from scaling up. Bridging all these issues, the question of assessing the effectiveness of ESD is key. For, although the role played by ESD in making SD a reality and achieving the SDGs is foreseen as positive, too little conclusive evidence has been brought to light to date, to the extent that the issue is now of critical importance. Moreover, it is specifically targeted as one of UNESCO’s seven key strategies for scaling up reform [19,20]: it will be necessary to “identify suitable, relevant and measurable indicators at every level—local, national, regional and international—and for each initiative and programme” [6].

The issue is even more pressing in higher education (HE). Indeed, in this highly competitive segment of the education system, much more than the adoption of national education reform plans or the latest UNESCO recommendations, rankings play a central role in the relative reputation of these institutions and their pedagogical choices. Yet these same rankings show every sign of belonging to the anti-ESD movement, with a strong weighting given to the economic performance of graduates who have not been taught the imperatives of ESD and the SDGs it serves. For example, in 2020, the Financial Times ranking factors relating to ESD accounted for only 2% of the total coefficients [21].

Does this limited focus on ESD of HE institutions and stakeholders such as ranking agencies signify a disregard for its importance or, on the other hand, is it due to the unavailability of solid evidence of the effectiveness of ESD and reliable ways to measure it?

The second could be true precisely because the challenges of assessing the impact and effectiveness of ESD are enormous and made up of both (1) a legitimacy issue, since it is a question of justifying the transition to ESD, through measurement and proof of its real impact, and of accelerating its systematic integration and mainstreaming (particularly within HE and its rankings) and the associated funding; and (2) an efficiency and performance issue, since it is a question of ensuring the identification and monitoring of good practices, key success factors and barriers to the implementation of ESD, in order to ensure both better learning and better teaching.

A first barrier to overcome will be how to measure the actual outcomes of ESD. Indeed, it would seem that traditional evaluation methods are not compatible with ESD in that they judge the acquisition of pure knowledge, while the holistic and pluralistic approach

of ESD aims at transmitting behaviors, values and ways of being compatible with SD. Additionally, the scope of traditional assessment methods is limited to what is acquired by the student whereas the assessment of ESD is meaningful only if the transition to concrete action (i.e., the translation of the values transmitted throughout schooling) is correctly monitored. In other words, while traditional methods seek to ensure the before/after control of knowledge, the evaluation of ESD looks for a gap between the before/after of sustainable practices and behaviors.

A second, and significant, barrier to overcome will be to assess whether these ESD outcomes are attributable to it and it alone. For, this is the main criticism addressed to studies aimed at evaluating the effectiveness of new practices: how can we certify that the results measured would not have occurred in any case, i.e., without ESD? In other words, how can the causal relationship between the inputs and outputs of these methods be highlighted by neutralizing all the elements and results that are not directly attributable to them?

More simply put: how can we assess the impact of ESD, and based on what results?

1.4. Towards a Better Understanding of Doubts around ESD Impact

For more than 30 years since the defining moment of the Brundtland Report (1987) [3], efforts towards greater sustainability have proliferated. In response to these challenges, ESD is now considered a leading tool for a more sustainable world. Behind this concept, enshrined by UNESCO, lies the simple imperative of equipping young citizens with the skills, knowledge and values needed for a world more compatible with SD, notably through a holistic, pluralistic and student-centered approach.

Today, while commitments to ESD are common, these new pedagogies are still anecdotal and difficult to scale up. Instead of the systemic transformation that it calls for, ESD is often barely relegated to a supplementary role in an education system that is still everywhere anchored in outdated paradigms.

At the heart of these difficulties lies, among other things, a lack of legitimacy of ESD, a direct consequence of the difficult measurement of its effectiveness. Although there is little doubt about the positive impact that ESD generates on people and the planet, empirical references clearly linking ESD to its benefits are still rare. In a global context of the “managerization” of educational practices, which must increasingly demonstrate their transparency and accountability, the stakes are all the more pressing. Among the effectiveness measurement methods employed, only “impact” assessment methodologies offer a reliable overview of ESD’s capacity to train sustainable citizens: they alone make it possible to highlight the gap in attitudes and skills between the trajectory of a group that has benefited from ESD and the hypothetical trajectory of the same group without ESD.

In the absence of similar pre-existing work, the aim of this literature review is to foster scaled ESD research and initiatives by offering a better understanding of the doubts surrounding the “impact” (stricto sensu) of ESD. At the crossroads of econometrics, educational sciences and psychology, this paper (1) sheds light on the methods and good practices for assessing this impact; (2) underlines the specificity of the data to be collected in the context of these methods of assessment; and (3) outlines the existing conclusions of impact studies dedicated to ESD that have served to highlight the limits and challenges for accurate measurement. In addition to its theoretical interest for the research community by presenting a compilation of impact assessment methods and evidence of the effectiveness of ESD, this paper is also intended to have a more practical ambition and is aimed at any practitioner interested in knowing the key success factors of an ESD approach and the relevant methods for measuring its effectiveness. Our research question is as follows: what impact assessment methods are relevant when it comes to the impact of ESD and what can we learn from their applications, challenges and results?

The remainder of this paper is set out as follows: we will first introduce the methodology used to frame the synthesis of the literature. Inspired by social science impact assessment methodologies, we will then prove that it is indeed possible to identify ESD outcomes that go beyond absolute effectiveness and that allow us to certify the causal character between inputs and outputs outside of any other contingencies or biases. We will then immerse ourselves in the panel of studies assessing the impact of ESD and their difficult collection of field data aimed at measuring not the knowledge but the skills of students. A third and final stage will investigate the results of these studies and their pedagogical implications.

2. Materials and Methods

In order to answer our research questions and test our hypothesis that the impact of integrating ESD into education programs is not only measurable but also translates into a concrete contribution to SD in the professional and personal practices of former ESD students, our study will use two main sources.

2.1. Systematic Literature Review

The first tool used in this study was a systematic literature review of ESD impact assessment case studies as a means of establishing a reliable evidence base. However, and in contrast to similar studies, notably those undertaken by O'Flaherty and Liddy [22] or Ardoin, Nicole and Bowers [23], only studies using strict quasi-experimental impact assessment methods were selected, in the hope of demonstrating with certainty the causal relationship between ESD and ex post sustainable behaviors. The aim was therefore to focus on works proving the effectiveness of ESD using control or treatment group comparisons, with a description of the interventions, clear objectives and conclusions. Mechanically, our review therefore focused on academic works or case studies presented in official reports of local, national and international organizations for a heterogeneous audience of relatively young learners (<25 years old). Similarly, and in view of the rapidly changing status quo in SD and ESD, only studies of a relatively recent nature were included (2000–2020). Finally, the studies collected are in English, French or Spanish, in full knowledge of the limitations generated by our team's language barrier. Our literature search strategy was essentially electronic and targeted databases related to Education (ERIC, MDPI, etc.) or more generic web catalogues. Keywords used were: "impact", "education for sustainable development", "environmental education", "sustainable citizens", "effectiveness ESD" and other variations of these terms. After a first screening of $n = 187$ contents, $n = 83$ articles retained our attention for further reading. Finally, and in view of our criteria above, $n = 21$ papers constituted our working base.

2.2. Interviews

To complete the panorama provided by the theory, two interviews were carried out in the framework of this study in order to bring an alternative and additional light anchored in the concrete practices of ESD. Due to the additional, more consultative nature of these interviews, only two were conducted, with two French experts on the subject: Anne Monnier, in charge of ecological transition in the schools of the ITM network; and Aude Serrano, co-founder of Impact Campus and independent consultant for the French organization Enseignants de la Transition.

The two 40 min sessions took the form of semi-directive interviews of five questions each on ESD practices and the assessment of its effectiveness in the higher education institutions of which the two interviewees were the direct referees. The content of these interviews served as a practical and expert-oriented perspective to select and challenge content and articles. The content of these interviews served as a practical and expert-oriented perspective to select and challenge content and articles.

3. Assessing the Impact of Educational Practices, General Framework for Analysis and Good Practice

3.1. Defining Impact Assessment

In its broadest sense, Maas and Liket [24] define impact as all the effects generated by an organization or intervention of any kind (public policy, program, product, technology, etc.) felt outside that same organization in society and on its environment, even at a distance. Narrowing the scope, the British organization Research Excellence Framework [25] defines the impact of education specifically as “an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia”.

Koehn and Uitto [26] provide another definition, which comes closer to the one that will guide most of the studies in our literature review: they will therefore be qualified as impact all “real-world changes in ecological sustainability, policies, and people’s well-being” specifically made possible by the acquisition of SD competencies, whether they are:

- Positive for society (we speak of positive externality), or negative;
- Direct, or indirect;
- Immediate or staggered in time [27].

The impact assessment will therefore aim to highlight this concrete change when it is directly attributable to a program, practice or, to use the scientific term, “treatment”. In concrete terms, it will be a question of highlighting a difference in trajectory and results between a situation that has been exposed to a specific “treatment” and the same situation should nothing have changed. In this sense, impact assessment is to be differentiated from simple evaluation/assessment tools, which mainly seek to express a change in absolute value, without neutralizing the many factors that suggest that part of the final result would have occurred no matter what. In other words, and in our case, the aim is to reveal the part of individual, programmatic or collective changes exclusively attributable to ESD and not to any other cause, in order to highlight the causal nature between education and the generation of positive impact. But how can this discrepancy be concretely accounted for? How can one simulate these two situations to be compared? This is the aim of impact assessment methods, which are part of a long econometric tradition.

3.2. Principles of Impact Assessment

The methodology of “theory-based impact evaluation” [28], which is today the dominant approach in the social and educational (for the sake of clarity, a clear distinction should be made between the general framework of analysis of theory-based impact evaluation and the impact assessment “models” it contains, which refer directly to the methodologies used in the formation of interest groups) sciences, proposes a process built around six imperatives, which constitute as many interconnected stages-precepts:

1. Mapping the results chain based on the “theory of change”;
2. Understanding the general context;
3. Anticipating heterogeneity;
4. Applying impact assessment methods to help form groups whose trajectories will be compared;
5. Applying a rigorous factual analysis method;
6. Favoring mixed methods of data collection.

More synthetically, a distinction must be made between: a preliminary stage enabling the bases of the measure to be laid (these are the first three items mentioned above), the choice and deployment of the impact assessment methodology (items 4 and 5) and, finally, the choice and analysis of the corresponding data (item 6) [29].

3.2.1. Preliminary Steps of Impact Assessment

The first step in an impact assessment study will aim to reconstruct the expected causal pathway of an intervention, starting with the expected final impact(s) of ESD on

students, then working backwards through the pattern of effects and causes to the inputs and learning activities that are meant to set the overall process in motion. This intellectual exercise, known as the “theory of change” [30–32], should be carried out with all the program’s stakeholders, both internal and external, in order to obtain a clear and common “results chain” (the result chain is defined as the concrete deliverable of the theory of change), taking into account all the issues at stake and making it possible to reveal the conditions for change.

Starting from the formulation of an initial problem and in parallel with a survey of the elements of context, influences and heterogeneity in the data to be collected, the results chain will therefore make it possible to map and reveal the causality between:

- Inputs, i.e., all the financial, human and educational resources for the project;
- Activities, i.e., the set of educational interventions responsible for transforming inputs into outputs;
- Outputs, i.e., the results of the program visible in behavioral change “on paper”;
- Final outcomes, i.e., the concrete results visible in behavioral change applied “in real life” and visible at the individual, programmatic or collective level.

“Theory of change” and “results chain” will therefore be preliminary steps to clarify objectives, risks and possibly reveal intermediate targets and impacts, since each element added to the overall scheme will have to be accompanied by specific, measurable, realistic and targeted indicators [29]. Other elements are then essential both from the point of view of statistical validity (question of the duration of the study, or the size of the treatment group and its antagonist, the comparison group) and from the administrative point of view (in particular the study budget).

With the aim of revealing concretely the link between the effects and causes highlighted in this first stage, the decision on an evaluation question that will guide the impact assessment will be the second stage of the process. This question will need to be precisely formulated in order to test the hypotheses underlying the “theory of change”: for example, while most studies generally question the generic effectiveness of a measure or program that is targeted through impact assessment (“What is the impact of an ESD program on behavioral change? On the acquisition of specific SD skills?” or others), one should keep in mind that the question can also take more original forms (“What is the impact of such an ESD program on chewing gum consumption?”, for example). Let us note that it will always be a matter of questioning the causal nature of the change at work: such studies will therefore always assess the impact of something on something else to be defined. (The axiom is particularly important since it prescribes any idea of measurement that is far too generic. Typically measuring “the overall impact of ESD” or even “the impact of ESD on the achievement of SDGs” seems too broad to be dealt with adequately. See in particular the methodological details below.)

Other questions that should guide the measure and its relevance include interrogations about the potential of the study [29]. Is it:

- Innovative? Will the measure used make it possible to highlight a new, untested approach that has yet to prove its worth?
- Replicable? Will the design of the assessment tool be able to identify contextual similarities that will allow the program being tested to be scaled up? (A distinction is generally made between two types of impact measurement: efficacy studies which are established in an extremely precise and controlled context and whose results are often too specific to have any value in the context of scaling up a practice or program, and effectiveness studies which are established in a more “normal”, less regulated context and which therefore have a greater value of replicability).
- Strategic? What will be the strategic significance of the measure for the survival of the program or its reform?

While the research question will need to remain largely unchanged throughout the impact assessment process, the theory of change may need to be modified to accommodate

surprises in the data: the aim will be to ensure that it is the original model (and therefore the conclusions reached) that fits the data, and not the other way around.

3.2.2. Choosing an Impact Assessment Method

When it comes to evaluating the impact of any educational (or social) measure, simply noting some kind of result at the end of the day is not enough. Let us imagine an evaluation that aims to show the link between ESD and sustainable behavior, and that actually proves in the end an increase in students' behavior in this direction: how can we know that these results are directly attributed to the benefits of ESD?

Much more than a result expressed in absolute values, it is therefore the comparison with a basic hypothetical trajectory that will make it possible to validate the exclusive causal link between a cause and its effect. In other words, the impact arises from the comparison between the results of an actual situation that has been influenced by a program and a hypothetical situation that is assumed to have remained unchanged.

How can we model this hypothetical situation, create these two identical "starting" "units" (the unit refers to the target audience of the study: are they individuals, households, countries, etc.?) in order to simultaneously observe the evolution of their trajectories under the effect or not of the program? Because if it is, in the end, easy to obtain a measurement of the results of the unit subjected to the program, and in the absence of the possibility of cloning the starting unit or of using a time machine to change the basic parameters, how can we obtain the results of this same unit this time not subjected to the basic treatment?

This is the whole object of counterfactual analysis [29], which from the outset will seek two separate units that are so similar that they can be assimilated to each other. So much so that, in the end, it will be possible to estimate almost certainly that, because of a strong initial similarity, the trajectory followed by the unit that did not follow the treatment is similar to the hypothetical trajectory followed by the unit that did follow the treatment, even if it would not have done so.

In concrete terms, to find these units and their mirrors, and as it is statistically impossible to find enough common characteristics of only two units that will be compared back to back, impact assessment methods use the power of numbers to form two groups of units that will share, on average, over a large number of units on each side, the same characteristics. Called "counterfactual analysis", this approach will therefore create two groups: a "comparison group" acting as a statistically identical duplicate of the "treatment group". Four imperatives stand out for the formation of the treatment group and the counterfactual estimate [29] (the estimation of the counterfactual here refers directly to the formation of the comparison group, a statistical mirror of the treatment group):

1. The characteristics of the two groups "on average" must be substantially equal in the absence of a program;
2. The results obtained "on average" by the two groups should be the same if they both experienced the treatment;
3. The treatment must never affect the comparison group in a direct or indirect way (risk of "spillover effect");
4. The treatment must imperatively be followed by all the members of the treatment group and by none of the members of the comparison group (risk of "imperfect compliance").

By ensuring that these four conditions are met, it will be certain that the variation observed on average in the treatment group will be solely attributable to the program, since this will be the only difference to which the two groups, which were initially identical from a statistical point of view, will be subject.

So how exactly does one form the two groups? Here, several models are in competition and can be classified according to whether they belong to one of these three approaches: experimental, quasi-experimental or non-experimental.

The experimental approach, so called because it is the one that corresponds exactly to the experience formulated by the theory, is represented by a single model, the “Randomized Controlled Trial” (RCT), which measures the average difference in results between two statistically identical groups formed randomly. When randomization occurs on a large enough pool of potential participants and without any other intervention, the only difference between the two groups formed will be the difference in treatment received or not received. At the present time, and because of its closeness to theoretical prescriptions, RCT is still seen as the flagship method of impact measurement [33], provided that it meets the conditions for success, particularly in terms of the number of units participating in the initial pool, which must be sufficient to form the two statistical clones. Other difficulties are also pointed out [34]: in particular, the concern of cognitive biases at work that are not taken into account by the model, the logistical burden, the follow-up difficulties, the lack of conformity, etc.

In contrast to the experimental approach and RCT, quasi-experimental methods move away from the theoretical and randomized framework by allowing evaluation teams to intervene in the formation of groups, mainly in a bid to map and neutralize biases at work. These include [29]:

- Selection bias, where performance is likely to be correlated with the desire to participate in the program or not;
- Lack of compliance, including the appearance in a given group of defectors from the opposing group;
- The Hawthorne effect, which occurs when individuals increase their performance simply because they feel they are the object of special attention;
- The John Henry effect, which exposes the comparison group to increased performance to compensate for not receiving treatment;
- Substitution bias, which occurs when some units in the comparison group that are disappointed about not receiving treatment voluntarily expose themselves to substitute treatments that bring their performance closer to that of the treatment group.

Because of these biases, which must be neutralized, the differences measured between groups by quasi-experimental methods often cannot be exploited as such, but will be returned in the form of local estimates that can be interpreted at the level of sub-groups and which must be analyzed carefully. Among these quasi-experimental approaches are the methods of [29]:

- Matching, which aims at individually selecting the units of the treatment group and looking for an identical individual counterpart that shares the exact same characteristics of interest and that will constitute a mirror unit in the comparison group.
- “Difference in differences” (double difference method, which will be discussed extensively here), which aim to compare the results of the comparison and treatment group before and after (unlike other methods that only consider the “photo finish”), on the assumption that biases and unobservable factors do not vary over time and that if the initial measurement takes them into account for both groups, the final measurement will also do so, thus neutralizing them.
- Stepped wedge design, which aims to convert a treatment group that is supposed to receive treatment at a later time into a comparison group in the present time, in order to circumvent selection bias that might have led some individuals (or units) to accept the experience simply to receive treatment. The underlying assumption is

that if the group was intended to become a future treatment group, it has then all the characteristics required to be a good comparison group as well.

- Instrumental variables, which form the basis of a systematic and factor-based recruitment strategy for both groups (the so-called instrumental variables) directly correlated with selection but without affecting the results.

Non-experimental approaches are called so because they do not “create” a control experiment: they do not base their results on the measure of a gap in the outcomes of a comparison/treatment group. Rather, they assess the results of the treatment group alone at different points in the program (“interrupted time series”) or simply before and after the program (“pre/post test”). While RCT is perceived as the most reliable method if properly implemented, the non-experimental approaches, on the other hand, are undoubtedly the ones offering the least guarantees. Indeed, the postulate of the two tools mentioned above is cavalier: it suggests that between the “pre” and the “post” or between the different intermediate snapshots, no other factor than the simple fact of undergoing the treatment will have influenced the group’s result. If the hypothesis seems valid in static, highly predictable environments and over very short periods of time, it is directly questioned when one of these elements is not present. In this case, the measure is no longer valid: it is impossible to determine whether the impact generated really comes from the program or from other causes.

The choice of the most relevant method when it comes to assessing the impact of ESD needs to be made carefully. Several criteria will have to be taken into account, in particular, and it will be necessary to favor the method which:

- Best matches the operational context of the program: the resources available, the quality and quantity of potentially eligible units, the ease of the research team with a given method;
- Necessitates the formulation of as few conditions and assumptions as possible in order to validate the counterfactual estimate;
- Requires the least amount of data.

On these last two items and in increasing order of difficulty of implementation, let us mention the RCT, then the set of quasi-experimental methods, with matching in last position. In case of doubt, however, the possibility of using several methods to triangulate the results is commonly accepted. In some cases, impact assessment methods are also complemented by other, more participatory approaches, which are closer to simple but less reliable measures of impact (one can think in particular of the Most Significant Change methodology [35]).

3.2.3. Data Collection and Analysis of Results

The final stage, which involves collecting data to feed into the indicators defined at the outset, will very often be based on representative samples defined within the treatment and comparison groups themselves. The data collected will depend directly on the question guiding the research and may therefore be relative to:

- The outcomes of the experience: this is the majority of data collected empirically.
- The experience’s intermediate results: these are the data linked to the intermediate indicators defined during the mapping of the theory of change, and which allow the validation or invalidation of the relevance of the expected causal chain.
- The activities: this is the data which, in our case in particular, allows us to highlight good teaching practices.

The impact measure could be:

- Prospective, i.e., aiming to measure results based on data produced at the end of the course. This type of method has in particular the advantage of: (1) allowing a collection of pre-intervention data favorable to the creation of relevant comparison groups, (2) contributing to the creation of ambitious objectives for the program to experience the

potential for impact measurement at the end of the course as an additional incentive to achieve them.

- Retrospective, i.e., using data already collected to highlight a change in the past. This type of analysis is to be preferred in order to measure the impact of past programs, with a view to possible comparison with new current developments; however, the method does not generally allow for reliable comparison groups, which directly threatens the reliability of the measurements.

They may be of a nature that is:

- Qualitative: questionnaires, interviews, observation, case studies, videos, photos, etc. Qualitative data have a descriptive scope which allows us to adapt to variations in context, to account for nuances in depth and to highlight subtle dysfunctions.
- Quantitative: surveys, test results, anthropometric measurements, etc. These have the advantage of complementing the contribution of qualitative tools by offering a more direct, robust and interpretable general lighting.

Specifically targeted by the theory-based impact evaluation approach, the use of mixed methods will be favored. Researchers [36–38] highlight some advantages of using qualitative and quantitative tools:

- Triangulation of results, increasing their robustness and credibility;
- Use of results of one type to build more efficient tools of another type during the course of the study;
- Complementarity of the results which expose different facets of the same phenomenon.

In practice, this topic of collecting the data needed for ESD impact assessment will be key.

4. Assessing the Impact of ESD, from Theory to Practice

Emerging in the wake of early environmental concerns, ESD is a type of education aimed at creating “sustainability citizens” [5] capable of meeting the human, environmental and economic challenges of today and tomorrow. While ESD is widely acclaimed in the literature and is being implemented worldwide, the new practice is still struggling to scale up.

Behind this delay seems to be an efficiency that has not yet been fully proven. In order to overcome this initial obstacle, and beyond simple evaluations showing ESD results in absolute terms, more robust measures to assess its impact have been multiplying over the last 10 years or so. It will now be a question of being able to show the strict superiority of ESD over traditional methods in terms of its capacity to train the citizens of today and tomorrow.

At this stage, therefore, the most common question addressed in ESD impact assessments is to prove its transformative character compared with a more traditional educational scheme. In short, is ESD really succeeding in achieving its goals of shaping responsible citizens?

From the definition of this question of mainstream impact research, a first element of difficulty emerges: if the measurement of impact is in the collection of data reporting on the achievement of ESD pedagogical objectives, what should these objectives and their associated results ultimately be? Additionally, and to take the implications of the theory-based impact evaluation framework further, how can we move from the pedagogical objectives to the data that can be used by an impact evaluation model?

4.1. Identifying the Pedagogical Objectives of ESD

Let us start by addressing the first question: what would be the pedagogical objectives of ESD that a potential treatment group could achieve in order to account for its impact?

In the mid-1990s, when the concept of ESD was still in its infancy, the question of its educational objectives was primarily analyzed using the “information deficit model” [39,40], which makes knowledge acquisition the simple determinants of behavioral change in favor of SD. In other words, this rational model believed that simply educating for SD would directly promote the acquisition of certain sensitivities that would trigger pro-environmental behavior.

Since then, however, this theory has come under strong criticism. For, if a lack of knowledge often prevents change, the acquisition of knowledge is not enough (this focus on knowledge and content, however, unfortunately remains central to the design of most programs [41]). Indeed, mere “declarative” knowledge (that relating to facts and data) is not sufficient to promote change, which can only be achieved through the transmission of three additional types of knowledge, namely [42–44]:

- Procedural knowledge: a set of “know-how” acquired in the field through direct contact with the issues taught [13];
- Effectiveness knowledge: the set of perceptions that make action desirable and that emerge naturally from practice and debate [13];
- Social knowledge: all information about the aims and intentions of others [13], as well as their perception in the form of norms.

Thus, and beyond simple declarative knowledge (i.e., the cognitive elements of knowledge) ESD will also aim to transmit other elements of an affective nature, this time relating to the know-how and skills specific to SD, which alone will be capable of initiating the transformation of behavior in the long term. From a pedagogical and educational point of view, and at all levels, it is therefore a major shift that is at work: that from an education for knowledge to an education for action and competence.

In the sphere of education, Kliem and Leutner [45] define “competence” precisely as those “context-specific cognitive dispositions that are acquired and required in order to cope with certain situations or tasks in specific areas”. First introduced in linguistics [46], then in educational sciences [47], it is psychology that definitively endorses the concept. Already at the time, McClelland [48] called for education to be able to “assess skills rather than intelligence”. In fact, unlike intelligence, a moving notion that allows us to account for generic individual capacities independent of any context, competence reflects the individual potential for cognitive and affective response to a specific request or situation: in this sense, the term is close to what is expected in “real life”; it is, as Connell, Sheridan, and Gardner [49] point out, “realized abilities”.

Adopted by UNESCO, and echoing the Bologna Process that started in 1999 which allowed the standardization of a predominantly European university system based on common competences [50], the approach endorses an emancipatory vision of ESD, which Vare and Scott [51] and Wals [52] call “ESD 2”, to help build the capacity to think and act critically in tomorrow’s world (in opposition to what these same authors call “ESD1”, a more normative vision where ESD directly promotes certain modes of behavior instead of aiming to develop the capacity to act with autonomy of thought). The aim will be to be able to define the expected competences in order to see the pedagogical objectives of ESD become a reality: to enable students to become agents of change capable of dealing with systemic, ambiguous, uncertain, changing problems, and to become managers and leaders in the transition towards more SD and in all sectors.

Since this founding moment, many researchers have therefore endeavored to draw up a list of the skills needed to establish the three pillars of SD and their connections. A non-exhaustive list of existing reference frameworks and terminologies include the following:

- The “sustainable skills” of Wals [52];
- The “Gestaltungskompetenz” (“Shaping competencies”) of de Haan [53];
- Glasser and Hirsh’s “Core Competencies for Sustainability” [54];
- Rieckmann’s “key competencies” [55];
- The “key competencies for sustainability” of Wiek et al. [56,57];

- The OECD's "key competencies", defined in the Definition and Selection of Competencies (DeSeCo) project [58];
- Stibbe's "sustainability literacy" [59].

Despite some visible differences in detail, and without exhaustively citing all the references on the matter, however, these different frameworks provide a list of key competencies that provide a consensus on what is expected as core SD ways of acting that would reflect the intrinsic qualities of ESD. For example, UNESCO's summary list of competencies includes [6]:

- System thinking competency, i.e., "the ability to recognize and understand relationships, to analyze complex systems, to perceive the ways in which systems are embedded within different domains and different scales, and to deal with uncertainty";
- Anticipatory competency, i.e., "the ability to understand and evaluate multiple futures—possible, probable and desirable—and to create one's own visions for the future, to apply the precautionary principle, to assess the consequences of actions, and to deal with risks and changes.";
- Normative competency, i.e., "the ability to understand and reflect on the norms and values that underlie one's actions and to negotiate sustainability values, principles, goals and targets, in a context of conflicts of interests and trade-offs, uncertain knowledge and contradictions";
- Strategic competency, i.e., "the ability to collectively develop and implement innovative actions that further sustainability at the local level and further afield";
- Collaboration competency, i.e., "the ability to learn from others; understand and respect the needs, perspectives and actions of others (empathy); understand, relate to and be sensitive to others (empathic leadership), deal with conflicts in a group; and facilitate collaborative and participatory problem-solving";
- Critical thinking competency, i.e., "the ability to question norms, practices and opinions; reflect on one's values, perceptions and actions; and take a position in the sustainability discourse";
- Self awareness competency, i.e., "the ability to reflect on one's own role in the local community and (global) society, continually evaluate and further motivate one's actions, and deal with one's feelings and desires";
- Integrated problem-solving competency, i.e., "the overarching ability to apply different problem-solving frameworks to complex sustainability problems and develop viable, inclusive and equitable solutions that promote sustainable development—integrating the above-mentioned competencies".

Taken together, these skills and ways of being form what Axelrod and Lehman [60] call "sustainable behavior", sometimes also called "sustainable literacy", i.e., all the behaviors, knowledge and actions that contribute to SD. They will need to be developed specifically through ESD and, from the perspective of higher education, specifically target the jobs for which they prepare students [61], together with the more basic set of skills that are specific to the role of education in its broadest sense [56].

Complementing this first effort, the work of the German research group Transfer 21 has led to a more granular result thanks to the establishment of 12 sub-competencies, themselves subdivided into finer elements, and supplemented by suggestions for suitable teaching content and approaches [62].

Capitalizing on this work, and in the wake of the establishment of the SDGs with which ESD is directly involved, UNESCO published a report [63] in 2017 specifically detailing the pedagogical objectives in the form of "cognitive", "socio-emotional" and "behavioral" skills specific to the achievement of each of the SDGs, as well as the suggested contents and approaches. In other words, and completing the Transfer 21 project, UNESCO has detailed the SD competencies by transposing them into the framework of the SDGs and thereby produced a list of expected outcomes of ESD, which form a solid basis for ESD impact measures.

This brings us to the question of how we can move from these pedagogical objectives to the data that can be used by an impact evaluation model.

4.2. Relevant Data for the Impact of ESD

In a literature review dating from 2017 and 2018, Ardoin, Nicole and Bowers [23] and O’Flaherty and Liddy [22] undertook an initial effort to collect all the studies aimed at demonstrating the effectiveness of ESD, resulting in a corpus of respectively 119 and 34 publications (sometimes overlapping) that brought together: (1) actual impact assessment work, with a comparison group; and (2) more classical work of measuring results without a comparison group.

By eliminating the second category of studies, whose methods do not make it possible to unequivocally highlight the results directly attributable to ESD (although they are useful since they can triangulate some results of the first type of analysis), and by adding more recent work, the literature offers some 21 studies [64–84] evaluating the impact of ESD in terms of the achievement of its objectives. These studies vary by:

- The sample size of the starting pool.
- The duration of data collection.
- Geographical location (mainly Western and Northern Europe, the United States and Turkey).
- The type of learner targeted.
- The evaluated attribute: the studies measure the attainment of one or more skills (study on recycling behaviors in Redman [76]), one or more types of knowledge (everywhere, study based on the comparative analysis of cognitive and affective gains), one or more teaching practices (study on the influence of role-playing in Paschall and Wüstenhagen [75]), etc. As already mentioned, it should be noted that no study aims to directly measure the contribution of ESD to the whole spectrum of SDGs. Indeed, according to Redman [76], global approaches generally represent a logistical challenge related to the size and power of the tools to be developed. In the absence of such rigour, the results are generally assumed to be too weak to reach any conclusions.

Depending on these parameters, some studies are therefore mechanically more robust than others, or at least more generalizable in their conclusions. Nevertheless, they all share an adherence to quasi-experimental methods of impact assessment, with a treatment and comparison group established directly by the researchers. For most of them, also and more precisely, the approach will be similar to the double difference method, which aims to compare the performance gap between the two groups before and after the introduction of treatment (it is therefore an inter- and intra-group difference). Among these surveys, about a quarter of the analysis focuses on the difference in results between the profile of “traditional” students and those integrated into “eco-schools”, primary and secondary schools which are labelled according to the school’s performance in its pedagogical approach to SD.

The subject of collecting data useful for these studies directly raises the question of methods for assessing skills, the central corpus of the new pedagogies.

Should we therefore turn our attention upstream, and report on the simple acquisition “on paper” of these SD skills, at the price of a heavy wager on the probable concretization of these skills, which may as well not find the elements of their realization in “real life” and therefore have no concrete consequences for the SDGs?

Or, on the contrary, should we place our attention further downstream in the process by reporting only on sustainable behaviors that are actually actualized, realized, and translated into real life (as an example of this translation of skills into reality, let us cite in particular the behaviors towards more exemplary citizenship (sorting and recycling, energy savings, a reduction in meat consumption, carpooling, train instead of plane, etc.) and professional exemplarity (activism, sustainable leadership) for which it is estimated that aggregation at the collective level can bring about SD), in other words, measuring the “performance” (performance is defined as a realized competence (Shohamy [85])) behind

the acquisition of skills? Indeed, the work of the behavioral sciences quite rightly warns that mastering a skill “on paper” will never certify with certainty its automatic translation into real life. Without going into the details of the extremely fertile theories of behavioral sciences, let us illustrate the facts by the work of Blake [86] in particular, who identifies three barriers to action (the field covered is that of the “attitude–action gap”):

- Individuality, which means that an individual will always put his or her “selfish” desires first before taking action;
- Responsibility, which consists of saying that an individual will only act if he feels that his action will serve a purpose, or if he feels that it is his responsibility to act, which models of pro-environmental behavior call the “locus of control” [87–90]; and
- Practicality, which describes the societal and institutional barriers, apart from any intention or attitude, that prevent action (e.g., lack of time, money, and difficulty in accessing the right information).

This dichotomy of approaches to measuring competence is not new. In 1990, in a study on hospital competence, Miller [91] already called for a distinction to be made between several levels of concretization of individual competence, from the lowest level of mere procedural knowledge to the perfect mastery of the competence achieved in the concrete passage to action.

Thus, just like Miller, who proposes different evaluation tools depending on whether the measure focuses on the “Knows/Knows How/Shows How” (the “theoretical” competence) or on the “Does” (the “realized” competence, the performance) as highlighted in his pyramid of competency mastery, our impact studies, which cover both notions, should also use different instruments to measure competencies according to their level of ambition in terms of restitution of ESD results.

All the studies covered by our review (whether their ambition is to measure the “Knows/Knows How/Shows How” or the “Does”) thus give priority to the use of primarily quantitative tools of a positivist nature, as Rickinson [92] has already noted: the aim is to use tests to collect a quantified measure of the competence acquired through ESD. Here, two approaches can be distinguished in the selection of tools. Some studies use the following approaches:

- Mainstream and general tools developed independently of the context and sometimes adapted to the study: we can mention in particular the EDINSOST tools [82], PISA [81], the New Ecological Paradigm scale of sustainable behavior [71], which directly echo other reference frameworks not used in these studies—notably, the Sulitest, the Global Environmental Behavior Scale, the Environmental Literacy Questionnaire, the Environmental Literacy Survey, etc.;
- Tools built from scratch for the good of the assessment (in Redman [76] in particular), the theoretical merit of this approach being that it offers a more accurate measure of the competence assessed “in context”.

In both cases, the approaches used will be the same: the aim will be to draw up a list of competences (and possibly sub-competences) with several levels of mastery targeted by the approach and to build a measurement model based on tools which comply with the rules of validity and reliability of psychometrics.

Even if there is no dominant model in the literature regulating these quantitative tools, the studied approaches seem to have common characteristics, in particular that of recreating real-life situations in order to provide as accurate a measure as possible of the estimation of the cognitive response in context. Most of the time, the tool will be structured as follows:

- Presentation of a starting scenario close to a critical situation of reality, in the form of text, image, film, graphics, etc.;
- Questions of various forms designed to inform the candidate's cognitive skills in context.

On this second point, it is worth noting the predominance of two practices which allow the results to be triangulated in order to get closer to psychometric standards:

- Open-ended questions are particularly relevant in this case since they require students to put together a coherent assembly of abstract, multidisciplinary elements contributing to the resolution of the problem highlighted in the scenario. The grading system will usually include a clear qualitative guide for markers on the expected performance standards [93–96].
- Closed-ended questions (sometimes presented as MCQs) have the advantage of being objective, relevant and automatically highlighting the mastery of a skill, since the student is directly called upon to eliminate incorrect answers and reveal the one that seems right to him or her. For a lower processing cost [97], this method makes it possible to estimate the competence in a way that is at least equivalent to open-ended responses [98].

For example, students may have to choose the most relevant source among several types of information related to an SD issue, assess the validity of an element based on concrete evidence, summarize or explain an issue, choose among different summaries the one that best suits a situation, etc.

In this respect, research increasingly seems to favor computer-based test formats [99] which favor complex stimuli that are more valid and close to reality (notably via a wide variety of scenario formats) [100], interactive procedures, multiple response formats, real-time adaptation of items to skills already demonstrated earlier in the test [101], and the appearance of immediate feedback [102]. However, it should be noted that such tests should not be motivated by simple numerical development but should always be based on valid theoretical and psychometric foundations.

Using quantitative tools that provide an initial exploitable basis, most impact studies focusing on the measurement of the transcription in real life of the skill (the “Does” level of mastery) generally complement the measurement by means of two devices:

- A module for monitoring the skills acquired over the long term, sometimes with tests carried out up to several months after the end of the experience [69];
- A student-centered interpretative qualitative approach aimed at detecting in the field the indicators of a behavioral inflection, and in particular: observation, interviews, analysis of student logbooks, etc. The aim is to move from the assessment of competence in a simulated context to an assessment established in a real context.

If there is one tool common to both approaches to measuring competence (a rather quantitative approach to the estimation of the “Shows How” and at the same time quantitative/qualitative approach to the estimation of the “Does”) and which is the subject of much controversy, it is the self-declarative questionnaire, which invites participants to fill in some of their attitudes, convictions, knowledge and motivations themselves. The tool is not valid from a psychometric point of view as it is highly exposed to numerous biases, in particular the desirability bias, which consists of candidates modifying their answers in favor of behaviors that are supposedly more desirable than others. (An original approach aimed at retaining the self-reporting format while neutralizing the desirability bias can be found in Hansmann, Mieg and Frischknecht [103] in their surveys sent to former graduates of a Swiss university with the aim of assessing their sustainable behavior. In addition to the self-declarative multiple choice questions, space is provided for respondents to document “concrete examples” of the reported facts.)

Overall, however, and in order to triangulate the results, both the literature and empirical evaluations tend to favor multi-method approaches, underlining the difficulty of measuring the achievement of ESD objectives.

So what results can be expected from these studies?

5. What Are the Results and Implications for ESD Impact Assessments?

Overall, most impact studies aimed at proving the effectiveness of ESD point to encouraging results in terms of its ability to impart the knowledge and skills of tomorrow compared with more traditional modes of education.

Murray, Goodhew and Murray [71], in their study to demonstrate the impact of ESD approaches centered around the sustainable values and behaviors of 67 students over eight months, find that their qualitative tools have led to a greater environmental sensitivity of the exit learners compared with the control group. Four months after the end of the final teaching modules, several students testified to an “awareness that SD represents a state of mind rather than separate tasks”. It is therefore a change of global perspective that is at work.

Seeberg and Minick [78], who examine the effectiveness of cross-cultural competency approaches in the acquisition of global skills and perspectives among a population of 23–25-year-old aspiring teachers over a four-year period, also note greater open-mindedness, a reconsidering of preconceptions, a positive effect on communication skills, and increased sensitivity to others and to the power of community, all of which are SD-specific competencies as defined by numerous benchmarks.

Through an eight-week ESD course leading to a United Nations environmental negotiation simulation, Paschall and Wüstenhagen [75] also observe clear results in terms of the acquisition of cognitive and affective knowledge of their students, manifested in their ability, at the end of the course, to better understand global warming and its challenges, to integrate the impact of global warming on the economy and vice versa, and to solve complex problems related to the environment.

In his impact study comparing the environmental “attitudes, knowledge, habits and concerns” of students from eco-school-certified primary schools (and therefore supposed to convey ESD) and students from a conventional school, Ozsoy [73] found that the subjects from the former group performed better on all items. For example, after treatment, where the treatment group cited “pollution by greenhouse gas emissions”, “car emissions”, “industrial pollution”, “toxic waste”, “unsanitary conditions”, “the hole in the ozone layer” and “global warming” as personal issues of major concern, the control group cited only “toxic waste” and “the hole in the ozone layer”.

Redman [76], in a study to measure the impact of an ESD intervention on the recycling and eating behaviors of six summer program students, notes an increase in declarative, procedural, efficacy and social knowledge, as well as behavior change between the beginning and end of treatment. In a sub-sample of three students followed over the long term (one year), Redman also observed a relative maintenance of the new practices acquired during the program (for example, Jane and Jill’s progress in never using reusable bags when shopping prior to treatment, and now using them systematically).

The common denominator of these studies is the results, which are sometimes highly correlated with socio-demographic variables and which can provide a benchmark of good ESD practices.

On the one hand, on the question of socio-demographic variables, certain studies reveal results that are notably conditioned by gender [69,74]. Indeed, the impact of ESD on women is generally greater than on men. These findings corroborate similar findings [104,105] which argued that while men tend to have more declarative knowledge about the environment than women, women are distinguished by more emotional commitment to SD. This is finally a further confirmation of the importance of emotional knowledge in the change needed for SD.

On the other hand, and much more significantly, these studies also make it possible to draw up a list of examples of good pedagogical practice as required by the pluralistic approach to the concept. In this sense, this work responds to the injunctions of UNESCO, which has been asking since the ESD Decade to shed light, not only on inputs and outputs,

but also on the process, on the pedagogical methods at work in the black box. One example is the approaches of:

- Seedberg and Minick [78], who report on the use of an essentially dematerialized format for debate between stakeholders from all over the world;
- Goralnik, Habron and Thorp [81], who describe an intervention format including lectures, field experience and active student participation in the design of interventions, particularly with regard to evaluations;
- Paschall and Wüstenhagen [75], whose intervention structure follows a classic lecture format during the first five weeks, and a “student-led” format during the last two weeks, leading to a role-playing game featuring the learners in a United Nations environmental negotiation simulation;
- Ozsoy [73], which reports on a curriculum where, in addition to benefiting from an SD approach in each of their subjects, students directly use their skills by doing small environmental projects in their school;
- Cincera and Krajhanzl [68], who describe a program centered around small diagnostic presentations on campus sustainability by students who are then required to design an action plan and monitor and evaluate progress towards improvement;
- Redman [76] is the most prolific in describing the approaches used, which are always plural and student centered.

By highlighting these new pedagogical practices, these studies therefore corroborate the results of previous work which testify to a socio-constructivist approach to education where attention is no longer focused (1) on teaching but on the pupil; (2) on the inputs but on the outputs; and (3) on the content but on the resolution of complex problems.

The new pedagogies of ESD, highlighted by these studies and the literature, will therefore advocate multi-method, experiential, active approaches in order to facilitate cognitive but above all affective learning. Among these new pedagogical approaches [106], some of which are described in our review, we can mention the following in particular:

- Concrete learning situations that place students in real-life situations in a context close to their own, helping them to reuse declarative knowledge acquired elsewhere and to develop their affective skills, testing their ability to solve complex problems, interact with a community, and question their values and representations [107–111]. Sterling [112] notes that this kind of approach allows “young people to gain confidence and a belief that they can make a difference, and their efforts can stimulate action by parents and the broader community”. Traditionally, concrete learning situations can be broken down into four approaches: inviting the concrete situation in the classroom (e.g., inviting speakers), going out to meet the concrete situation (e.g., organizing visits, trips), simulating the concrete situation (typical example of the role-play proposed by Paschall and Wüstenhagen [75], and questioning the concrete situation (e.g., a student in the role of a journalist questioning members of his or her household).
- Critical problem-solving, which begins its pedagogy by presenting a complex problem that is not supposed to have only one solution [113], and which pushes students to a collective, experiential, non-normative resolution, supposed to help them form their own idea of the issues under discussion, question their relationship to the world and to others [114], and encourage the consideration of multiple points of view and approaches and the emergence of a concrete action plan involving the whole class. The aim will be to find tangible problems that resonate with the context and experience of pupils, readable on several scales and with varying degrees of difficulty. At this stage, the teaching posture will be that of a facilitator and a co-learner, capable of promoting the emancipation of pupils.
- Active learning that places the student at the center of the process and forces him or her to take part in the very design of pedagogical approaches.

Other studies showing mixed or even weak results of ESD include Boeve-de Pauw and Van Petegem [66], Boeve-de Pauw, Gericke, Olsson and Berglund [79], Berglund, Gericke and Chang Rundgren [64], Hallfredsdottir [70], Krnel and Naglic [84] all report better performance in the cognitive elements of ESD of students from eco-schools in Flanders, Sweden, Iceland and Slovenia compared with students from conventional curricula, without any real change in the behavior of the former.

In addition to underlining the importance of pedagogical approaches and thus the pluralistic aspect of ESD for change, these studies have the advantage of highlighting some of the difficulties associated with the implementation of ESD by teachers accustomed to more traditional methods. In particular, the researchers of these teams mention:

- Teachers who are shy or even foreign to ESD in their approach, with methods that give pride of place to the “true/false” and “correct/incorrect” normative debate, which is incompatible with these issues;
- Teachers alien to ESD in its content: several studies, notably at Berglund, Gericke and Chang Rundgren [64] underline the difficulty of recognizing the whole spectrum of SD, with efforts mostly focused on the environmental pillar. The phenomenon is part of an archaic SD tradition where environmental issues were a central value.

More generally, these two points, which illustrate the difficulties of holism and pluralism in their practical implementation, echo more substantial criticisms of both predicates. We can cite here the work of Kopnina [9] who identifies two limits to the virtues of pluralism, notably in that it will never be fully democratic since the discourse on SD is dominated by corporatist and economic perspectives, and also in that it only conveys anthropocentric perspectives by rejecting at the margin the perspectives of non-human animals in particular.

It is therefore a new role that ESD proposes to the teacher, since by exposing him or her to the need to adapt his or her methods and content to the major challenges of the 20th century, ESD pushes the teacher to the need to have a “vision” of the issues at stake in SD, in an epistemological imperative which is that of pedagogy as a deliberate criticism of modes of operation to be banished.

While this is a common denominator in most studies showing little impact for ESD, there are other elements that explain some of the differences in performance within certain groups at times:

- Olsson, Gericke and Chang Rundgren [74], and Uitto and Saloranta [83] note particularly insignificant results among one treatment group, pointing in particular to the “lack of a clear understanding of the role of the treatment in the treatment process” and lack of motivation (motivation is defined by Preuss [115] as “a desire to concrete action”) of the pupils in the group;
- Murray, Goodhew and Murray [71] point to unaccounted-for biases as well as a tool design based on a pool of insufficiently significant participants (67 in all), which brings us back to econometric considerations.

Finally, and more generally, whether they foresee a positive or negative impact, all these studies should not overlook a large number of psychosocial factors which call for caution in the interpretation of such results, including the following: the impact of specific cultures [116], childhood values [42], emotional involvement [117], ... conscious and unconscious, specific and generic parameters that act as stimulators or inhibitors of pro-environmental change [44] and that reveal the complexity of measuring the impact of ESD.

6. Conclusions

This literature review has allowed us to provide a deeper look into impact assessment methods relevant for demonstrating the impact of ESD, as well as their applications, challenges and results.

While studies assessing the impact of ESD share similarities with other impact assessment studies, special attention needs to be drawn to the use of “theory-based impact evaluation” [23] in the specific context of ESD. First, the causal pathway must be well

defined while the related question seems more likely to be innovative, replicable and strategic if it is precise and original. Second, the complexity of the task may require the combination of several methods to triangulate the results or even the addition of more participatory approaches.

The challenges of ESD impact assessment methods are multifaceted and include the difficulty to follow the principles of counterfactual analyses [29], various biases [71] as well as the isolation of the effects of ESD from a large number of psychosocial factors [42,116,117].

Results found in studies demonstrate that ESD has brought about encouraging outcomes in students, including greater environmental sensitivity [71], a reconsidering of preconceptions [78], an improved ability to solve complex problems related to the environment [75], a greater likelihood of naming environmental issues as personal concerns [73] and a relative maintenance of the new positive practices acquired [76].

As a condition for the future expansion of ESD, the question of assessment and proof of its effectiveness is a central body of research around the concept. Of the many approaches explored, (quasi-) experimental impact measurement appears to be the most viable in that it only reveals the outcomes of ESD when they are solely and directly attributable to it. It is precisely these outcomes, unlike traditional education, that will target competence, i.e., the readiness for action oriented towards the three pillars of SD and that will need to be skillfully measured.

Navigating through these difficulties, the impact studies of some 20 research teams are mostly optimistic about the capacity of ESD to transform individual perspectives towards SD and the realization of the SDGs. Beyond the many biases and psychosocial factors at the origin of different degrees of behavioral change, ESD can only be achieved by making its holism, and above all the pluralism of its approach, even more concrete and in ways that allow ecocentrist perspectives to blossom.

These studies, focusing mainly on small groups and being difficult to compare, combined with other studies' more hesitant conclusions, can explain some of the doubts surrounding the impact of ESD and its current limited scale. However, the logistical challenge presented by global approaches [69] may play a part in the limited scale of many of these studies.

In order to confirm the more positive results, the literature agrees on the need for long-term longitudinal impact studies, taking into account other types of concrete ESD results that can be realized in a sometimes more distant horizon than what current studies can cover (activism in particular).

Another area for future ESD impact assessments to invest in would be the more specific area of higher education. Indeed, can the results of the work undertaken so far primarily in the context of primary education effectively be generalized to the context of higher education? In other words, would there not be certain specificities of higher education that would make the conditions for the success of ESD different in this very particular context? This remains to be proven.

Finally, Bourn [118] warns against the tendency to too often predefine the expected outcomes of ESD. For such a new concept, he advocates remaining open to unexpected surprises that can be read in the data and that are capable of reorienting the implementation strategies of the reform.

Author Contributions: Conceptualization, Q.S.; methodology, Q.S.; formal analysis, Q.S.; investigation, Q.S.; resources, Q.S.; writing—original draft preparation, Q.S.; writing—review and editing, Q.S., J.W. and C.H.; supervision, J.W. and C.H.; project administration, Q.S., J.W. and C.H.; funding acquisition, J.W. and C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by La Belle EDuC, an impact-first company (registered as a Société à Mission in France in 2020) dedicated to accelerating the integration of sustainability content in higher education teaching material and the student learning experience.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: This review was funded by La Belle EDuC with the aim of furthering research in the area sustainability education, a mission to which it (as a French Société à Mission) is legally bound through its statutes, which include “contributing to the development and sharing of measurement tools and best practices for education stakeholders” and “raising awareness among stakeholders of the challenges of sustainability and its integration in teaching and the student learning experience”. The authors declare no conflict of interest.

Author Biographies: Quentin Ssossé graduated in 2020 with the highest honors from prestigious French schools HEC Paris and Ecole Polytechnique. During his Master’s of Research in Management, he specialized in Innovation Management. His research work has led him to assess the sustainable transition strategies of large corporate groups directly on the field: while working with a leading industrial multinational, he focused particularly on questions of environmental labeling by participating in the co-construction of a multi-criteria impact measurement framework. Firmly convinced that the environmental transition cannot take place without education, Quentin joined the Working Group of La Belle EDuC to help make education for sustainable development mainstream. Johanna Wagner is a consultant and university lecturer with a background in asset management and hotel operations. She is passionate about people, sustainability and innovation, which has led her to teach in leading hospitality master’s programs in Europe including IMHI/ESSEC and ESG Business School. She holds an MBA from IMHI/ESSEC (2013) and a certificate in Sustainability Reporting and CSR Strategy from ESSEC/Deloitte. She is Co-Founder of La Belle EDuC, an impact-first company (registered as a Société à Mission in France) dedicated to accelerating the integration of sustainability content in higher education teaching material and the student learning experience. Carina Hopper is a sustainability lecturer teaching at the master’s level at ESSEC Business School, SKEMA Business School and ESMOD Fashion Business School, among others. She is also active in the start-up community, advising and collaborating with tech and social start-ups across Europe. She holds an MBA from ESSEC Business School (2018) and a certificate in Business Sustainability Management from the University of Cambridge. She is Co-Founder of La Belle EDuC, an impact-first company (registered as a Société à Mission in France) dedicated to accelerating the integration of sustainability content in higher education teaching material and the student learning experience.

References

1. Fressoz, J.B. *L’Apocalypse Joyeuse Une Histoire du Risque Technologique*; Seuil: Paris, France, 2012.
2. Meadows, D.H. *The Limits to Growth: A Report for the Club of Rome’s Project on the Predicament of Mankind*; Universe Books: New York, NY, USA, 1972.
3. World Commission on Environment and Development (WCED). *Our Common Future*; Oxford University Press: Oxford, UK; New York, NY, USA, 1987.
4. Shukla, P.R.; Skea, J.; Buendia, E.C.; Masson-Delmotte, V.; Pörtner, H.-O.; Roberts, D.C.; Zhai, P.; Slade, R.; Connors, S.; van Diemen, R.; et al. *IPCC: Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*; IPCC: Rome, Italy, 2019; in press.
5. Wals, A.E.J.; Lenglet, F. Sustainability citizens: Collaborative and disruptive social learning. In *Sustainability Citizenship in Cities: Theory and Practice*; Horne, R., Fien, J., Beza, B., Nelson, A., Eds.; Routledge: London, UK, 2016.
6. UNESCO. *Issues and Trends of Education for Sustainable Development*; UNESCO: Paris, France, 2018; pp. 1–68.
7. Orr, D. What is Education for? *Context* **1991**, *27*, 52–59.
8. Stapp, W.B.; Bennett, D.; Bryan, W.; Fulton, J.; MacGregor, J.; Nowak, P.; Swan, J.; Wall, R. The concept of environmental education. *J. Environ. Educ.* **1969**, *1*, 30–31. [[CrossRef](#)]
9. Kopnina, H. Education for Sustainable Development (ESD): The Turn Away from “Environment” in Environmental Education? *Environ. Educ. Res.* **2012**, *18*, 699–717. [[CrossRef](#)]
10. Hopkins, C. Twenty Years of Education for Sustainable Development. *J. Educ. Sustain. Dev.* **2012**, *6*, 1–4. [[CrossRef](#)]
11. Öhman, J. *Values and Democracy in Education for Sustainable Development*; Liber: Malmö, Sweden, 2008.
12. Wals, A.E.; Corcoran, P.B. Re-orienting, re-connecting and re-imagining: Learning-based responses to the challenge of (un)sustainability. In *Learning for Sustainability in Times of Accelerating Change*; Wals, A.E.J., Corcoran, P.B., Eds.; Wageningen Academic Publishers: Wageningen, The Netherlands, 2012; pp. 21–32.
13. Wals, A.E. *Social Learning towards a Sustainable World: Principles, Perspectives, and Praxis*; Wageningen Academic Publishing: Wageningen, The Netherlands, 2009.
14. Cortese, A. The Critical Role of Higher Education in Creating a Sustainable Future. *Plan. High. Educ.* **2003**, *31*, 15–22.
15. Foxman, D.; Bateson, G. Steps to an Ecology of Mind. *West. Political Q.* **1973**, *26*, 345. [[CrossRef](#)]
16. Leal, F.W. About the role of universities and their contribution to sustainable development. *High. Educ. Policy* **2011**, *24*, 427–438.
17. Leal, F.W. Future Challenges for Sustainable Development, 2012. University World News. Available online: <http://www.universityworldnews.com/article.php?story=20120613184239690> (accessed on 17 June 2012).

18. Leal, F.W.; Manolas, E.; Pace, P. Education for sustainable development: Current discourses and practices and their relevance to technology education. *Int. J. Technol. Des. Educ.* **2009**, *19*, 149–165. [[CrossRef](#)]
19. Ofei-Manu, P.; Didham, R.J. *Learning Performance Assessment in Education for Sustainable Development: Investigating the Key Factors in Effective Educational Practice and Outcomes for Sustainable Development*; UNU-IAS and IGES: Hayama, Japan, 2012.
20. EFA Global Monitoring Report. *Education for All 2000–2015: Achievements and Challenges*; UNESCO: Paris, France, 2015.
21. Kirkland, D.; Buchan, E. Integrating Sustainability into Business Education: Part II (Online). Business Graduates Association, 2020. Available online: <https://businessgraduatesassociation.com/integrating-sustainability-into-business-education-part-ii/> (accessed on 28 February 2021).
22. O’Flaherty, J.; Liddy, M. The impact of development education and education for sustainable development interventions: A synthesis of the research. *Environ. Educ. Res.* **2017**, *24*, 1031–1049. [[CrossRef](#)]
23. Ardoin, N.; Bowers, A.; Roth, N.; Holthuis, N. Environmental education and K-12 student outcomes: A review and analysis of research. *J. Environ. Educ.* **2017**, *49*, 1–17. [[CrossRef](#)]
24. Maas, K.; Liket, K. Social impact measurement: Classification of methods. In *Environmental Management Accounting and Supply Chain Management*; Springer: Berlin, Germany, 2011; pp. 171–202.
25. HEFCE. *Research Excellence Framework: The Results*; Higher Education Funding Council for England: London, UK, 2014.
26. Koehn, P.H.; Uitto, J.I. *Universities and the Sustainable Development Future Evaluating Higher-Education Contributions to the 2030 Agenda*; Earthscan: London, UK, 2017.
27. Lebeau, Y.; Cochrane, A. Rethinking the ‘third mission’: UK universities and regional engagement in challenging times. *Eur. J. High. Educ.* **2015**, *5*, 250–263. [[CrossRef](#)]
28. White, H. Theory-based impact evaluation: Principles and practice. *J. Dev. Eff.* **2009**, *1*, 271–284. [[CrossRef](#)]
29. Gertler, P.J.; Martinez, S.; Premand, P.; Rawlings, L.B.; Vermeersch, C.M.J. *Impact Evaluation in Practice*, 2nd ed.; Inter-American Development Bank and World Bank: Washington, DC, USA, 2016. [[CrossRef](#)]
30. Weiss, C.H. How Can Theory-Based Evaluation Make Greater Headway? *Eval. Rev.* **1997**, *21*, 501–524. [[CrossRef](#)]
31. Weiss, C.H. Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*; Connell, J., Ed.; Aspen Institute: Washington, DC, USA, 1995.
32. Weiss, C.H. *Evaluation Research: Methods of Assessing Program Effectiveness*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1972.
33. Rossi, P.H.; Lipsey, M.W.; Freeman, H.E. *Evaluation: A Systematic Approach*; Sage: Thousand Oaks, CA, USA, 2004.
34. Bamberger, M.; White, H. Using Strong Evaluation Designs in Developing Countries: Experience and Challenges. *J. Multidiscip. Eval.* **2007**, *4*, 58–73.
35. Dart, J.; Davies, R. Dialogical, Story-Based Evaluation Tool: The Most Significant Change Technique. *Am. J. Eval.* **2003**, *24*, 137–155. [[CrossRef](#)]
36. Booth, D. Towards a better combination of the quantitative and the qualitative: Some design issues from Pakistan’s participatory poverty assessment. In *Q-Squared—Qualitative and Quantitative Poverty Appraisal: Complementarities, Tensions and the Way Forward*; Kanbur, R., Ed.; Department of Applied Economics and Management Cornell University: Ithaca, NY, USA, 2005; pp. 59–63.
37. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*; Sage: Thousand Oaks, CA, USA, 2007.
38. Greene, J.C. Synthesis: A reprise on mixing methods. In *Discovering Successful Pathways in Children’s Development: Mixed Methods in the Study of Childhood and Family Life*; Weisner, T.S., Ed.; The University of Chicago Press: Chicago, IL, USA, 2005; pp. 405–419.
39. Nisbet, M.; Scheufele, D. What’s next for science communication? Promising directions and lingering distractions. *Am. J. Bot.* **2009**, *1767–1778*. Available online: http://assets.bigthink.com/images/age_of_engagement/NisbetScheufele2009_ScienceCommunication_AJB.pdf (accessed on 18 September 2015). [[CrossRef](#)] [[PubMed](#)]
40. Dickson, D. The Case for a “Deficit Model” of Science Communication. 2005. Available online: <http://www.scidev.net/global/communication/editorials/the-case-for-a-deficit-model-of-science-communic.html>. (accessed on 28 September 2015).
41. Pooley, J.A.; O’Connor, M. Environmental education and attitudes: Emotions and beliefs are what is needed. *Environ. Behav.* **2000**, *32*, 711–723. [[CrossRef](#)]
42. Kaiser, F.G.; Woelfing, S.; Fuhrer, U. Environmental attitude and ecological behavior. *J. Environ. Psychol.* **1999**, *19*, 1–19. [[CrossRef](#)]
43. Kaiser, F.G.; Fuhrer, U. Ecological behavior’s dependency on different forms of knowledge. *Appl. Psychol.* **2003**, *52*, 598–613. [[CrossRef](#)]
44. Kollmuss, A.; Agyeman, J. Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* **2002**, *8*, 239–260. [[CrossRef](#)]
45. Klieme, E.; Leutner, D. Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms bei der DFG. *Z. Pädag.* **2006**, *52*, 876–903.
46. Chomsky, N. *Aspects of the Theory of Syntax*; MIT Press: Cambridge MA, USA, 1965; p. 4.
47. Roth, H. *Pädagogische Anthropologie. Band II: Entwicklung und Erziehung*; Schroedel: Hannover, Germany, 1971.
48. McClelland, D.C. Testing for competence rather than for intelligence. *Am. Psychol.* **1973**, *28*, 1–14. [[CrossRef](#)] [[PubMed](#)]
49. Connell, M.W.; Sheridan, K.; Gardner, H. On abilities and domains. In *The Psychology of Abilities, Competencies, and Expertise*; Sternberg, R.J., Grigorenko, E.L., Eds.; Cambridge University Press: Cambridge, UK, 2003; pp. 126–155. [[CrossRef](#)]

50. Shams, F.H.; Jeroen, C.C.; Hsieh, S.; Wilkins, A.C. The Bologna process and its impact in the European Higher Education Area and beyond. In *The SAGE Handbook of International Higher Education*; Deardoff, D.K., de Wit, H., Heyl, J.D., Adams, T., Eds.; Sage Publications: Thousand Oaks, CA, USA, 2012; pp. 81–100. [CrossRef]
51. Vare, P.; Scott, W. Learning for a change: Exploring the relationship between education and sustainable development. *J. Educ. Sustain. Dev.* **2007**, *1*, 191–198. [CrossRef]
52. Wals, A.E.J. *Beyond Unreasonable Doubt. Education and Learning for Socio-Ecological Sustainability in the Anthropocene*; Wageningen University: Wageningen, The Netherlands, 2015; Available online: https://arjenwals.files.wordpress.com/2016/02/8412100972_rvb_inauguratie-wals_oratieboekje_v02.pdf (accessed on 17 May 2017).
53. De Haan, G. The development of ESD-related competencies in supportive institutional frameworks. *Int. Rev. Educ.* **2010**, *56*, 315–328. [CrossRef]
54. Glasser, H.; Hirsh, J. Toward the development of robust learning for sustainability core competencies. *Sustain. J. Rec.* **2016**, *9*, 121–134. [CrossRef]
55. Rieckmann, M. Future-oriented higher education: Which key competencies should be fostered through university teaching and learning? *Futures* **2012**, *44*, 127–135. [CrossRef]
56. Wiek, A.; Withycombe, L.; Redman, C.L. Key competencies in sustainability: A reference framework for academic program development. *Sustain. Sci.* **2011**, *6*, 203–218. [CrossRef]
57. Wiek, A.; Bernstein, M.J.; Foley, R.W.; Cohen, M.; Forrest, N.; Kuzdas, C.; Kay, B.; Withycombe Keeler, L. Operationalising competencies in higher education for sustainable development. In *Routledge Handbook of Higher Education for Sustainable Development*; Barth, M., Michelsen, G., Thomas, I., Rieckmann, M., Eds.; Routledge: London, UK, 2016; pp. 241–260.
58. Rychen, D.S. Key competencies: Meeting important challenges in life. In *Key Competencies for a Successful Life and Well-Functioning Society*; Rychen, D.S., Salganik, L.H., Eds.; Hogrefe and Huber: Cambridge, MA, USA, 2003; pp. 63–107.
59. Stibbe, A. *Handbook of Sustainability Literacy: Skills for a Changing World*; Green Books: Cambridge, UK, 2010.
60. Axelrod, L.J.; Lehman, D.R. Responding to environmental concerns: What factors guide individual action? *J. Environ. Psychol.* **1993**, *13*, 149–159. [CrossRef]
61. Bertschy, F.; Künzli, C.; Lehmann, M. Teachers' Competencies for the Implementation of Educational Offers in the Field of Education for Sustainable Development. *Sustainability* **2013**, *5*, 5067–5080. [CrossRef]
62. De Haan, G. The BLK "21" programme in Germany: A "Gestaltungskompetenz"-based model for Education for Sustainable Development. *Environ. Educ. Res.* **2006**, *12*, 19–32. [CrossRef]
63. UNESCO. *Education for Sustainable Development Goals. Learning Objectives*; UNESCO: Paris, France, 2017. Available online: <http://unesdoc.unesco.org/images/0024/002474/247444e.pdf> (accessed on 1 December 2020).
64. Berglund, T.; Gericke, N.; Chang Rundgren, S.N. The Implementation of Education for Sustainable Development in Sweden: Investigating the Sustainability Consciousness among Upper Secondary Students. *Res. Sci. Technol. Educ.* **2014**, *32*, 318–339. [CrossRef]
65. Boeve-de Pauw, J.; van Petegem, P. The effect of Flemish eco-schools on student environmental knowledge, attitudes and affect. *Int. J. Sci. Educ.* **2011**, *33*, 1513–1538. [CrossRef]
66. Boeve-de Pauw, J.; van Petegem, P. The effect of eco-schools on children's environmental values and behavior. *J. Biol. Educ.* **2013**, *47*, 96–103. [CrossRef]
67. Boeve-de Pauw, J.; Jacobs, K.; van Petegem, P. Gender differences in environmental values: An issue of measurement? *J. Env. Psychol.* **2014**, *46*, 373–397. [CrossRef]
68. Cincera, J.; Krajhanzl, J. Eco-schools: What factor influence pupils' action competence for pro-environmental behaviour? *J. Clean. Prod.* **2013**, *61*, 117–121. [CrossRef]
69. Coertjens, L.; de Pauw, J.; de Maeyer, S.; van Petegem, P. Do schools make a difference in their students' environmental attitudes and awareness? Evidence from PISA 2006. *Int. J. Sci. Math. Educ.* **2010**, *8*, 497–522. [CrossRef]
70. Hallfredsdóttir, S. *Eco-Schools: Are They Really Better? Comparison of Environmental Knowledge, Attitudes and Actions between Students in Environmentally Certified Schools and Traditional Schools in Iceland*. Master's Thesis, Reykjavik University, Reykjavik, Iceland, 2011.
71. Murray, P.; Goodhew, J.; Murray, S. The Heart of ESD: Personally Engaging Learners with Sustainability. *Environ. Educ. Res.* **2013**, *20*, 718–734. [CrossRef]
72. Murray, P.; Douglas-Dunbar, A.; Murray, S. Evaluating Values-centred Pedagogies in Education for Sustainable Development. *Int. J. Sustain. Higher Educ.* **2014**, *15*, 314–329. [CrossRef]
73. Ozsoy, S.; Ertepinar, H.; Saglam, N. Can eco-schools improve elementary school students' environmental literacy levels? *Asia Pac. Forum Sci. Learn. Teach.* **2012**, *13*, 1–25.
74. Olsson, D.; Gericke, N.; Chang Rundgren, S.N. The Effect of Implementation of Education for Sustainable Development in Swedish Compulsory Schools—Assessing Pupils' Sustainability Consciousness. *Environ. Educ. Res.* **2015**, *22*, 176–202. [CrossRef]
75. Paschall, M.; Wuestenhagen, R. More than a Game: Learning about Climate Change through Role-Play. *J. Manag. Educ.* **2012**, *36*, 510–543. [CrossRef]
76. Redman, E. Advancing educational pedagogy for sustainability: Developing and implementing programs to transform behaviors. *Int. J. Env. Sci. Educ.* **2013**, *8*, 1–34.

77. Redman, E.; Redman, A. Transforming sustainable food and waste behaviors by realigning domains of knowledge in our education system. *J. Clean. Prod.* **2014**, *64*, 147–157. [[CrossRef](#)]
78. Seeberg, V.; Minick, T. Enhancing Cross-cultural Competence in Multicultural Teacher Education: Transformation in Global Learning. *Int. J. Multicult. Educ.* **2012**, *14*, 1–22. [[CrossRef](#)]
79. Pauw, J.B.-d.; Gericke, N.; Olsson, D.; Berglund, T. The Effectiveness of Education for Sustainable Development. *Sustainability* **2015**, *7*, 15693–15717. [[CrossRef](#)]
80. Chen, S.-Y.; Liu, S.-Y. Developing Students' Action Competence for a Sustainable Future: A Review of Educational Research. *Sustainability* **2020**, *12*, 1374. [[CrossRef](#)]
81. Habron, G.; Goralnik, L.; Thorp, L. Embracing the learning paradigm to foster systems thinking. *Int. J. Sustain. Higher Educ.* **2012**, *378–393*. [[CrossRef](#)]
82. Valderrama-Hernández, R.; Sánchez-Carracedo, F.; Alcántara Rubio, L.; Limón-Domínguez, D. Methodology to Analyze the Effectiveness of ESD in a Higher Degree in Education. A Case Study. *Sustainability* **2020**, *12*, 222. [[CrossRef](#)]
83. Uitto, A.; Saloranta, S. The relationship between secondary school students' environmental and human values, attitudes, interests and motivations. *Procedia Soc. Behav. Sci.* **2010**, *9*, 1866–1872. [[CrossRef](#)]
84. Krnel, D.; Naglic, S. Environmental literacy comparison between eco-schools and ordinary schools in Slovenia. *Sci. Educ. Int.* **2009**, *20*, 5–24.
85. Shohamy, E. Competence and performance in language testing. Performance and competence in second language acquisition. *Behav. Brain Sci.* **1996**, *19*, 138–151.
86. Blake, J. Overcoming the “value-action gap” in environmental policy: Tensions between national policy and local experience. *Local Environ.* **1999**, *4*, 257–278. [[CrossRef](#)]
87. Stern, P.C.; Dietz, T.; Kalof, L. Value orientations, gender, and environmental concern. *Environ. Behav.* **1993**, *25*, 322–348. [[CrossRef](#)]
88. Hines, J.M.; Hungerford, H.R.; Tomera, A.N. Analysis and Synthesis of Research on Responsible Environmental Behavior: A Meta-Analysis. *J. Environ. Educ.* **1987**, *18*, 1–8. [[CrossRef](#)]
89. Hungerford, H.R.; Volk, T.L. Changing learner behavior through environmental education. *J. Environ. Educ.* **1990**, *21*, 8–21. [[CrossRef](#)]
90. Sia, A.P.; Hungerford, H.R.; Tomera, A.N. Selected predictors of responsible environmental behavior: An analysis. *J. Environ. Educ.* **1985**, *17*, 31–40. [[CrossRef](#)]
91. Miller, G.E. The assessment of clinical skills/competence/performance. *Acad. Med.* **1990**, *65*, S63–S67. [[CrossRef](#)]
92. Rickinson, M. Learners and learning in environmental education: A critical review of the evidence. *Environ. Educ. Res.* **2001**, *7*, 207–320. [[CrossRef](#)]
93. Popham, W.J. What's Wrong and What's Right with Rubrics. *Educ. Leadersh.* **1997**, *55*, 72–75.
94. Osborn Popp, S.E.; Ryan, J.M.; Thompson, M.S. The critical role of anchor paper selection in writing assessment. *Appl. Meas. Educ.* **2009**, *22*, 255–271. [[CrossRef](#)]
95. Stellmack, M.A.; Konheim-Kalkstein, Y.L.; Manor, J.E.; Massey, A.R.; Schmitz, J.A.P. An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teach. Psychol.* **2009**, *36*, 102–107. [[CrossRef](#)]
96. Thaler, N.; Kazemi, E.; Huscher, C. Developing a rubric to assess student learning outcomes using a class assignment. *Teach. Psychol.* **2009**, *36*, 113–116. [[CrossRef](#)]
97. Thissen, D.; Steinberg, L.; Wainer, H. Detection of differential item functioning using the parameters of item response models. In *Differential Item Functioning*; Holland, P.W., Wainer, H., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1993; pp. 67–113.
98. Klein, H.; Cooper, J.; Molloy, J.; Swanson, J. The Assessment of Commitment: Advantages of a Unidimensional, Target-Free Approach. *J. Appl. Psychol.* **2012**, *99*, 222–238. [[CrossRef](#)] [[PubMed](#)]
99. Mayrath, M.C.; Clarke-Midura, J.; Robinson, D.H. *Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern*; Information Age Publishing: Charlotte, NC, USA, 2012.
100. Drasgow, F.; Schmitt, N. (Eds.) *Measuring and Analyzing Behavior in Organizations: Advances in Measurement and Data Analysis*; Pfeiffer: Zurich, Switzerland, 2002; Volume 10.
101. Linden, W.J.; van der Linden, W.J.; Glas, C.A. (Eds.) *Computerized Adaptive Testing: Theory and Practice*; Springer Science & Business Media: Berlin, Germany, 2000.
102. Chung, G.K.W.K.; O'Neil, H.F.; Bewley, W.L.; Baker, E.L. Computer-Based assessment in support of distance learning. In *Assessment of Competencies in Educational Contexts*; Hogrefe Publishing: Gottingen, Germany, 2008; pp. 253–276.
103. Hansmann, R.; Mieg, H.A.; Frischknecht, P. *Survey of Graduates of Environmental Sciences from 1992 to 2005: Analysis of their Careers and their Contributions to Sustainable Development*; DUWIS, ETH Zürich: Zurich, Switzerland, 2010.
104. Fliegenschnee, M.; Schelakovsky, M. *Environmental Psychology and Environmental Education: An Introduction of Human Ecological Perspective*; University of Vienna: Vienna, Austria, 1998.
105. Lehmann, J. *Befunde Empirischer Forschung zu Umweltbildung und Umweltbewusstsein*; Leske und Budrich: Opladen, Germany, 1999.
106. Andreotti, V. Soft versus Critical Global Citizenship Education. Policy and Practice: A Development. *Educ. Rev.* **2006**, *3*, 40–51.
107. Barab, S.A.; Luehmann, A.L. Building sustainable science curriculum: Acknowledging and accommodating local adaptation. *Sci. Educ.* **2003**, *87*, 454–467. [[CrossRef](#)]
108. Fadeeva, Z.; Mochizuki, Y.; Brundiers, K.; Wiek, A.; Redman, C.L. Real-world learning opportunities in sustainability: From classroom into the real world. *Int. J. Sustain. Higher Educ.* **2010**, *11*, 308–324. [[CrossRef](#)]

109. Frisk, E.; Larson, K.L. Educating for sustainability: Competencies & practices for transformative action. *J. Sustain. Educ.* **2011**, *2*, 1–20.
110. Segalàs, J.; Ferrer-Balas, D.; Mulder, K.F. What do engineering students learn in sustainability courses? The effect of the pedagogical approach. *J. Clean. Prod.* **2010**, *18*, 275–284. [[CrossRef](#)]
111. Steiner, G.; Posch, A. Higher education for sustainability by means of transdisciplinary case studies: An innovative approach for solving complex, real-world problems. *J. Clean. Prod.* **2006**, *14*, 877–890. [[CrossRef](#)]
112. Sterling, S. *Sustainable Education—Re-Visioning Learning and Change*; Green Books: Bristol, UK, 2001.
113. Hmelo-Silver, C. Problem-Based Learning: What and How Do Students Learn? *Educ. Psychol. Rev.* **2004**, *16*, 235–266. [[CrossRef](#)]
114. Tilbury, D.; Stevenson, R.B. (Eds.) *Education and Sustainability: Responding to the Global Challenge*; IUCN: Gland, Switzerland, 2002.
115. Preuss, S. *Umweltkatastrophe Mensch. Ueber Unsere Grenzen und Moeglichkeiten, Oekologisch Bewusst zu Handeln*; Roland Asanger Verlag: Heidelberg, Germany, 1991.
116. Meyer, L.; Sherman, L.; MaKinster, J. The effects of the Japan bridge project on third graders' cultural sensitivity. *Theory Res. Soc. Educ.* **2006**, *34*, 347–369. [[CrossRef](#)]
117. Chawla, L. Life paths into effective environmental action. *J. Environ. Educ.* **1999**, *31*, 15–26. [[CrossRef](#)]
118. Bourn, D. *The Theory and Practice of Development Education: A Pedagogy for Global Social Justice*; Routledge: London, UK, 2014.