Ecological Economics 201 (2022) 107546

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Is private sustainability governance a myth? Evaluating major sustainability certifications in primary production: A mixed methods meta-study

Thomas Dietz[a,*], Lisa Biber-Freudenberger[b], Laura Deal[c], Jan Börner[b]

[a] Institute of Political Science, University of Münster, Germany
[b] Center for Development Research, University of Bonn, Germany
[c] United Nations Environment Programme, USA

ARTICLE INFO

Keywords:
Voluntary Sustainability Standards (VSS)
Certification
Private governance
Effectiveness
Review
Meta-study

ABSTRACT

Sustainability certification (SC) is one of the most popular private sector approaches to govern social and environmental outcomes of trade in products from agriculture, forestry and fisheries. Based on a sample of 175 peer-reviewed articles, we use a novel mixed methods meta-analytical approach to study the success of major sustainability certifications in promoting sustainable (primary) production practices. We consider both qualitative and quantitative studies. Our main data source are the discussion and conclusion sections of research papers. We analyze conclusive statements about the success of SCs and categorize them into favorable, mixed, and skeptical evaluations. The picture is dominated by skeptical conclusions. Subsequently, we analyze how specific study characteristics affect this evaluation. The distribution of favorable, mixed, and skeptical evaluations is largely similar across the areas of economic, social, and environmental sustainability. Over time, the share of skeptical evaluations has increased. Contextual factors such as primary sub-sector, or country show no significant effects. The evaluations are also largely consistent across different types of SCs. Studies focusing on endpoint sustainability outcomes evaluate the performance of SCs significantly more skeptical than studies that focus on intermediate sustainability outcomes. Furthermore, our study shows that the share of skeptical evaluations significantly increases when a study examines the success of SCs for outcome variables with high implementation costs. Overall, our review points towards a limited success of SCs.

1. Introduction

Sustainability Certifications (SCs) such as the Forest Stewardship Council (FSC), Fairtrade (FT), or the Marine Stewardship Council (MSC) define rules for allegedly standard compliant markets and control access to these markets through certification (Cashore et al., 2004; Auld et al., 2009; Auld, 2014; Vogel, 2008). SCs feature most prominently in bio-based commodity chains in the three sub-sectors of (tropical) agriculture, forestry and fisheries/aquaculture (Tröster and Hiete, 2018). The main goal of SCs is to increase the economic, social and environmental sustainability of existing primary production systems. (Bartley, 2007; Cashore et al., 2004). As SCs continue to expand their geographic and market coverage, their performance at producer level has been increasingly studied (Oya et al., 2018; DeFries et al., 2017; Garrett et al., 2021; Dietz et al., 2020; Dietz et al., 2019).

Research on SC performance cuts across the social and natural sciences, including fields as diverse as economics, political science, anthropology, ecology and geography. Moreover, this literature has utilized a wide range of different quantitative and qualitative research designs to explore SCs' performance on the ground. Unsurprisingly, the results are mixed. Scholars have found both positive and negative sustainability outcomes of SC utilization, which in turn has spurred a fierce debate about the prospects of further upscaling SCs at producer level (Sellare et al., 2020)(Dietz et al., 2019; Sellare et al., 2020).

This article reviews a broad sample of this literature based on a mixed-methods strategy combining qualitative meta study approaches (Boyatzis, 2009) with descriptive and inferential statistics. In the discussion and conclusions sections scholars usually discuss and evaluate the findings that emerge from their empirical analysis. We develop a qualitative interpretive measure to synthesize these findings across the

[*] Corresponding author.
E-mail address: thomas.dietz@uni-muenster.de (T. Dietz).

https://doi.org/10.1016/j.ecolecon.2022.107546
Received 3 September 2021; Received in revised form 7 July 2022; Accepted 8 July 2022
Available online 5 August 2022

included studies. According to this analysis, conclusive statements about the on-site success of SCs fall into three broad overarching categories: favorable, mixed, and skeptical conclusions. We categorize the selected literature according to a pre-defined set of study characteristics (e.g. methods used, SCs included in the study, year of publication etc.) and quantify the distribution of favorable, mixed, and skeptical conclusions for each category. The combination of qualitative evaluation criteria with descriptive statistics allows us to recognize detailed patterns of how scholars evaluate the on-site success of SCs across a large body of heterogenous literature. We supplement this analysis with inferential statistical tests to better understand which of the patterns we identified in the descriptive analysis have already become more manifest than others.

We address the following research questions: First, how does the literature evaluate the success of SCs in improving the sustainability performance of certified primary production sites? Second, to what extent have these evaluations changed over time (*dimension of time*)? Third, to what extend do the evaluations vary across different SCs, different types of SCs, different areas of sustainability and different contexts (*dimension of scope*)? And, fourth, to what extent do we observe differences in the evaluations depending on whether an original study focuses on more or less ambitious economic, social or environmental sustainability outcomes (*dimensions of depth*)?

Our meta-study adds to the state of knowledge in various ways. Existing reviews are highly fragmented either in terms of the SCs they include or the areas of sustainability they focus on (Cattau et al., 2016; Carlson and Palmer, 2016; Bouslah et al., 2010; DeFries et al., 2017; Froese and Proelss, 2012; Parkes et al., 2010; Schleifer and Sun, 2020; Blackman and Rivera, 2011; Bray and Neilson, 2017; Terstappen et al., 2013; Garrett et al., 2021; Dammert and Mohan, 2015). Our study is the first that analyzes the success of all major SCs within the three subsectors of farm-agriculture, forestry, and fisheries/aquaculture across the areas of economic, social and environmental sustainability including the pertinent literature until 2020, and therefore presents the hitherto broadest account.

Moreover, the most advanced recently published meta-studies have used quantitative approaches to review the literature on the success of SCs at producer level (Oya et al., 2018; Meemken, 2020; Garrett et al., 2021). This involves calculating standardized numerical measures across original studies and calculated mean or average effect sizes for selected socioeconomic and environmental outcome variables. Through our focus on evaluative statements we provide a novel qualitative denominator to compare results across the pertinent literature. While quantitative meta-study approaches are certainly rigorous, the small number of original rigorous studies still limits their potential breadth and statistical power. The qualitative approach in turn, allows us to triangulate the literature across different research designs including quantitative and qualitative original studies and thus a richer and more inclusive literature base.

Further and most importantly, while effect sizes used in quantitative meta-studies represent a clear-cut measure to compare across studies, their informative value remains limited. The exclusive focus on numerical measures leaves out how authors of original articles interpret their empirical findings before they draw conclusions about the success of SCs. However, without this additional information it is often hard to understand whether an observed positive effect actually indicates a meaningful sustainability improvement or not. By synthesizing final qualitative statements about the success of SCs, our interpretative approach is well suited to capture the multiple dimensions and relevance of SC's sustainability effects.

In sum, the aim of this study is to enhance the knowledge base for both scholars and decision-makers in politics, economics and civil society concerning the now widely-discussed question of whether the problem-solving capacity of SCs is significant enough to drive or bolster the urgently-needed sustainability transformations that can safeguard global bio-based primary production in the current and coming decades. The article is organized as follows: First, we provide a short overview about the major SCs in the three bio-based sub-sectors of (tropical) farm-agriculture, forestry and fisheries/aquaculture, explain how they work and lay out the main economic, social and environmental sustainability outcomes they aim to achieve at producer level. Second, we explain our literature selection criteria and process, as well as the coding scheme we employ in our meta-study. Third, we display our results. Finally, we discuss our results and draw conclusions.

2. SCs in bio-based primary production

Almost three quarters (144) of the 203 private voluntary standard (VSS) schemes that the International Trade Centre has included into its comprehensive Standards Map are linked to bio-based primary production (https://sustainabilitymap.org/home). The vast majority of these voluntary sustainability standards are limited in scope either regarding the sustainability issues they cover or the proliferation rates they have achieved and have therefore little impact on sustainability outcomes. However, we also find a smaller group of well-established SCs, which have managed to grow into significant private governance actors (Auld, 2014; Dietz et al., 2018; Oya et al., 2018). As shown in Table 1 below, all major SCs that have emerged in primary bio-based production belong either to the sub-sector of (tropical) farm-agriculture, the forestry sector or the fisheries/aquaculture sector. Table 1 provides the details about this group of major SCs.

While these major SCs mentioned in Table 1 differ in various organizational details (e.g. number of commodities they cover, sustainability focus, strengths of standards, founding organizations etc.) (Dietz et al., 2018; Auld, 2014; Carlson and Palmer, 2016) it is notable, that at the same time they share a number of core features that define them as a specific governance mechanisms (Fransen and Kolk, 2007). At their most basic, all SCs have in common that they establish a set of economic, social and environmental sustainability standards that define how economic actors along global bio-based value chains ought to behave (Raynolds et al., 2007). SCs cannot mandate businesses to implement their rules, but depend on the rise of standard compliant markets pushing economic actors into their standard systems (Vogel, 2008). Under such conditions, value chain actors on all levels are expected to voluntarily agree to certification in order to access standard compliant markets (*standard adoption*) (Vogel, 2008; Grabs, 2020). After adoption SCs monitor compliance through third-party audits. Through decertification they can sanction defaulting parties and exclude them from standard compliant markets (Dietz et al., 2018). As such, SCs are intended to ensure that standard adoption induces value chain actors to implement standards and adapt their behavior accordingly (Carlson and

Table 1
Major SCs in bio-based production.

Sub-Sector	Sustainability Certifications
Forestry	• **FSC** (Forest Stewardship Council) • **PEFC** (Program for the Endorsement of Forest Certification) • **SFI** (Sustainable Forestry Initiative)
Farm-agriculture	• **RSPO** (Roundtable on Sustainable Palm Oil • **GlobalGAP** • **RFA** (Rainforest Alliance) • **UTZ** • **RTRS** (Roundtable on Responsible Soy) • **FT** (Fairtrade) • **FTorg** (Fairtrade/Organic) • **4C** • **BCI** (Better Cotton Initiative) • **CMA** (Cotton made in Africa) • **Bonsucro**
Fisheries/ aquaculture	• **GAA** (Global Aquaculture Alliance) • **FoS** (Friends of the Sea) • **MSC** (Marine Stewardship Council) • **ASC** (Aquaculture Stewardship Council)

Palmer, 2016; Cashore et al., 2004). Ultimately, these behavioral outputs are expected to improve economic, social and environmental outcomes at certified production sites.

In terms of economic sustainability, the main intended outcome of SC is to increase the income and well-being of certified producers in primary production and to reduce poverty especially in the Global South. Producers in developing countries are often located at the beginning of global value chains and provide primary goods for more advanced, knowledge intense and profitable intermediate production steps in developed countries. While lead firms from developed countries tend to capture the lion's share of the benefits of these business relations, producers from developing countries sometimes earn barely enough to cover production costs. In order to improve the economic conditions for certified producers, SCs demand downstream value chain actors to grant improved trading conditions to certified producers. These measures mainly include price premiums for standard compliant products but may also involve further mechanisms such as the granting of credits or longer-term contracts. Further, SCs may require certification holders to self-organize or offer specialized training on how to increase the productivity and profitability of their productions systems.

With respect to social sustainability, SCs aim to improve the working and living conditions of disempowered and marginalized groups such as dependent wage workers, women, children or indigenous people. Workers in many low-income countries around the world face conditions that violate basic human labor rights. The range of these legal violations ranges from unfairly garnished or withheld wages and unsafe or unhealthy working environments, to worst case scenarios, where production may involve child and slave labor. To improve the social sustainability in certified production sites, SCs usually include internationally accepted conventions on labor and human rights into their standard catalogues. Further, SCs require certified producers to engage in social investments, such as in schools, the establishment of grievance mechanisms, the provision of portable water and safe washing rooms, or the provision of protection wear for work in dangerous or unhealthy conditions. SCs may also include standards that promote the establishment of labor unions and worker representations in order to improve the social conditions for disempowered groups at producer level.

Finally, in terms of environmental sustainability, SCs strive to reconcile economic and social development with ecological sustainability goals. Bio-based production sites are often home to some of the most valuable ecological habitats on earth, including highly biodiverse rainforests or coastal waters. Increasing global demand and trade provides incentives to expand production into many of these habitats, which in turn may result in severe environmental damage and irreversible biodiversity losses. SCs address these issues by setting up environmental standards that regulate environmentally destructive activities. Examples for such standards include prohibitions to cut primary forest, prohibitions to catch fish beyond pre-defined catch quotas or bans of hazardous pesticides and fertilizers. In addition to that, SCs may oblige certified producers to invest in more environmentally friendly production practices, such as the use of ground cover to avoid soil erosion, the establishment of conservation zones to protect threatened species or the development of water management plans to safe water. All required investments into environmental practices follow the same goal of reducing the ecological footprint of bio-based primary production systems.

To be clear, not all SCs mentioned in Table 1 pursue exactly the same sustainability goals, nor do they use the same standards. However, even though SCs have their particularities, all major SCs have converged over time in the use of holistic and complex sustainability approaches integrating the three pillars of environmental, social and economic sustainability (Reinecke et al., 2012). The question is how successful these SCs indeed are in reaching the goals they claim to pursue.

3. Material and methods

3.1. Literature selection process

In order to identify relevant literature, we searched the scientific data base Scopus using a search string combining the names of all the major SCs outlined in Table 1 (see Supplementary material 1). With this strategy we aimed to identify all articles available via Scopus that deal with one or more of the major SCs previously selected. Since, to the best of our knowledge, there exists almost no scholarly literature concerning the success of SCs at producer level in bio-based primary production that does not address one of the major SCs included in Table 1 we deem our literature selection to be highly comprehensive. We deliberately excluded organic certification from our review. As shown by Meemken (2020), organic certification generates unique effects that systematically differ from the effects of other SCs and should therefore not be included into the same review. Meemken and Qaim (2018), and Seufert et al. (2012) provide specific reviews on organic standards. (See Table 2)

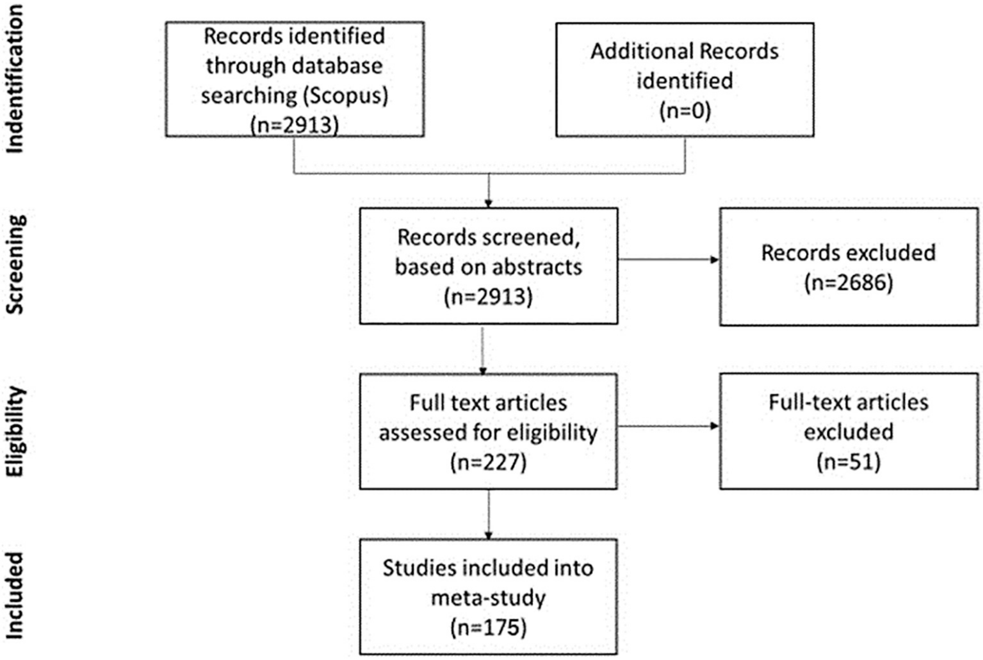
The initial search was performed on October 15, 2020, and yielded *2913* results. We used an inclusive approach to screen the abstract of the identified literature and selected all types of qualitative and quantitative research approaches that present and evaluate some kind of empirical evidence to study the social, economic or environmental performance of SCs on the ground of production. However, we ensured a high quality of the selected literature by accepting only articles that were credibly subject to processes of scientific peer review. Gray literature is often produced or contracted by organizations with specific interests in favor or against SCs, including the certifying bodies. There is always an underlying risk that this literature biased by conflicts-of-interests. In order to minimize the influence of such biases on our study, we thus excluded the gray literature from our literature base, and included only peer-reviewed studies published by scientific journals. Applying these criteria resulted in a sample of *227* records. In the third step, we assessed the eligibility based on the full-text articles. We excluded articles that did not refer to the sustainability success of SCs on the ground at producer level. This step further reduced the studies to a number of *175* records (see Supplementary material a).

The selected scholarly literature that studies the success of SCs at producer level can be roughly divided into three broad research approaches: First, (quasi-)experimental approaches, which statistically assess treatment effects by comparing certified producers to non-treated control groups. Second, quantitative observational approaches, which statistically study the effects of SCs on certified producers, however without comparing them to credible counterfactuals and third, qualitative approaches that use interpretative methods and in-depth case studies to evaluate the success SCs at producer level.

(Quasi-)experimental approaches are widely regarded as the most rigorous way to study the success of VSS on the ground because they control for selection bias (Baylis et al., 2016) Some meta-studies have used statistical techniques to synthesize the results of such studies. (Bray and Neilson, 2017; DeFries et al., 2017). While the attempt is certainly rigorous, it turns out, that the number of experimental studies remains too scarce to draw comprehensive statistical insights from their combined analysis (Oya et al., 2018; DeFries et al., 2017). Since both observational and qualitative approaches do not construct counterfactuals, they cannot directly analyze additionality, i.e., whether the effects they find are due to the impact of SCs or other external factors. Nevertheless, non-experimental approaches teach us a great deal about the state of sustainability in certified production sites. In line with Meemken (2020), we therefore believe that given the scarcity of rigorous (quasi-) experimental studies, combining both (quasi-)experimental and non-experimental studies will produce a more nuanced overall picture about the on-site success of SCs than analyzing each type of literature alone.

Table 2

Flow of information through different phases of the systematic review process *(n for number of records)*.


Indentification
Screening
Eligibility
Included
Records identified through database searching (Scopus) (n=2913)
Additional Records identified (n=0)
Records screened, based on abstracts (n=2913)
Records excluded (n=2686)
Full text articles assessed for eligibility (n=227)
Full-text articles excluded (n=51)
Studies included into meta-study (n=175)

3.2. Qualitative evaluation criteria

We use a qualitative interpretive approach to synthesize findings from the included studies. Since the literature included in our meta-study is highly heterogeneous, it lacks a simple common empirical denominator on the basis of which it could be easily compared. However, independent of the specific type of article, all articles in our literature base have in common that they draw conclusions about the success of SCs based on the contextualization and interpretations of their empirical findings. Such final evaluations are usually part of the discussion and/or conclusion section of an article. Our analysis started with an open coding process (Corbin and Strauss, 2015). We started to extract examples of conclusive statements from a randomly selected sub-sample of articles and synthesized these statements into overarching themes. Following the saturation concept by Bowen (2008), we continued this process until we found categories that adequately represent the universe of different potential evaluations. According to this analysis, conclusive statements about the on-site success of SCs fall into three broad overarching categories: favorable, mixed, and skeptical conclusions. Favorable conclusions present SCs as successful instruments by associating them with relevant sustainability improvements in the studied outcome variables. Skeptical views, on the other hand, consider SCs to be inappropriate instruments because they had either negative, no, or too weak effects on the outcome variables studied to generate relevant sustainability improvements. Mixed conclusions fall in-between these two categories, i.e. due to ambiguous empirical findings no clear insights about the on-site success of SCs can be drawn.

3.3. Coding scheme

We categorize the articles in our literature base according to their research designs. In terms of research approaches, we differentiate between quantitative (quasi-)experimental, quantitative observational, and qualitative research. In terms of data sources, we differentiated between the categories of direct observations (e.g. remote sensing, laboratory analysis, field observations, etc.), surveys (large n), in-depth interviews (small n), and documents. Overall, this differentiation allowed us to analyze the extent to which conclusions about the success of SCs depend on research design.

Next, we select a number of categories that enable us analyze the success of SCs at producer level along the three crucial dimensions of *time*, *scope* and *depth* (see introduction). To assess the time dimension, we group articles according to their year of publication. In doing so, we are able to see whether the distribution of favorable, mixed and skeptical evaluations has changed over time and the direction of current trends.

Subsequently, we add a set of seven categories to operationalize criteria of *scope*. First, we structured the articles according to the individual SCs they investigate (for the list of individual SCs see Table 1). Second, we grouped the individual SCs into two different types of SCs. All SCs we included in the study belong to so-called multi-stakeholder initiatives. Nevertheless, the role played by stakeholders from civil society (NGOs) and business (including business associations) within the organizational structure of different SCs varies greatly. Depending on what type of stakeholders dominates a particular SC, we distinguish between largely NGO/civil society-driven SCs and corporate/business association-driven SCs. Third, we differentiate the articles according to the outcome variables they focus on. Here, we distinguish on a relatively general level between environmental, social and economic outcome variables. Fourth, we structured the literature according to the subsector (farm-agriculture, forest, fisheries/aquaculture), and the concrete business sector (timber, coffee, seafood, palm oil, tea, cocoa, multiple/other) on which they focus. Sixth, we differentiated between smallholder and non-smallholder production sites. Seventh, we included country as a code also differentiating individual countries according to the strengths of relevant institutions based on various indices.

Finally, we add two further coding categories which allow us to operationalize the criterion of depth. First, for each article, we group the key stated outcome variables into intermediate or endpoint variables. Following Oya et al. (2018), we define intermediate outcome variables as a means to an end whereas endpoint variables reflect ultimate sustainability goals (e.g., price premiums to lift producers out of poverty, the use of ground cover to avoid soil erosion, the development of forest

Table 3
Coding scheme.

Coding area	Categories
Research Design	
Research approach	Quantitative (quasi)-experimental, quantitative observational, qualitative
Data sources	Direct observations, surveys (large n), in-depth interviews (small n), documents
Time Dimension	
Year of publication	Years 2000–2020
Scope Dimension	
Individual SCs	See Table 1.
Type of SCs	NGO/civil society-driven SCs, corporate/business association-driven SCs
Sustainability area of investigated outcome variable	Environmental, social or economic outcome variables
Primary sub-sector	Forest, farm-agriculture, fisheries/aquaculture
Business sector of certified production site	Timber, coffee, seafood, palm oil, tea, cocoa, multiple/other
Development status of certified production site	Smallholder, non-smallholder
Country of certified production site / differentiated according to institutional strength	All countries possible
Depth Dimension	
Intermediate vs. *endpoint outcomes*	Intermediate outcome variable, endpoint outcome variables
Implementation costs of investigated outcome variable	High, middle or low implementation costs

management plans to secure sustainable forests, the provision of protection wear to avoid unsafe work conditions). Naturally, success in the improvement of endpoint variables provides a much more rigorous measure to evaluate the success of SCs which is why we systematically differentiate these two types of outcome variables in our analysis.

Second, we distinguish between outcome variables with low, medium and high implementation costs. Following the categorization in Grabs (2020), we define sustainability outcome variables with low implementation costs as those for which there is a great deal of alignment between the business interests of the certified producers and the sustainability transformation goals are largely aligned (e.g. more efficient use of pesticides). Sustainability outcome variables with medium implementation costs follow a similar logic, but depend on costly investments in the present that only pay off in the future (e.g. change in production towards pest resistant crop varieties). Finally, we defined sustainability outcome variables with high implementation costs as outcome variables for which business interests of certified unit and sustainability goal contradict each other (e.g. paying minimum/living wages, banning the use of agrochemical inputs). Success in the improvement of outcomes with high implementation costs indicates deeper sustainable changes as compared to success in the improvement of outcomes with low or medium high implementation costs. The following table summarizes our coding scheme.

3.4. Statistical analysis

We apply this coding scheme to all articles in our literature base. We code 1 if a specific category is present in an article. Each time a category occurs in an article, it is considered as a separate case. We analyze for each case whether it is linked to a favorable, mixed, or skeptical evaluation. Take for example the coding area of individual SCs. Each SC included in Table 3 presents a separate category. If an article includes more than one SC, we analyze this article's evaluative statements separately for each individual SC. We group cases that include the same SC into sub-samples and use descriptive statistics to calculate the distribution of favorable, mixed, and skeptical evaluations per SC. We repeat this procedure for all categories in our coding scheme. (See Table 3)

We use the statistical software environment R (R Core Team, 2021) and the packages of tidyverse (Wickham, 2019; Wickham, 2019) and forcats (Wickham, 2021) to prepare the data for statistical analysis and visualization using ggplot2 (Wickham, 2016), cowplot (Wilke, 2020), ggpubr (Kassambara, 2020), and ggsn (Santos Oswaldo, 2019).

The combination of qualitative evaluation criteria with descriptive statistics allows us to recognize detailed patterns of how scholars evaluate the on-site success of SCs across a large body of heterogenous literature. This method therefore presents our major tool of analysis. We supplement this analysis with additional statistical tests: We use Pearson's Chi^2 Test in order to test whether the distribution of favorable, mixed, and skeptical evaluations between the categories within a coding area (see Table 3) becomes statistically significant. In those cases where we find statistically significant differences we complement the analysis with a post-hoc Chi^2 Test to identify which evaluation subset (skeptical, mixed, or favorable) is responsible for the significant difference (Ebbert, 2019). Additionally, we test the effect of country-specific secondary data such as the Human Development Index on the evaluation result using pairwise Wilcoxon Rank Sum Tests with Bonferroni correction. Finally, we use a binominal regression model to analyze the relative influence of categories specified in Table 3 on how the literature evaluates the success of SCs. Given the heterogeneity across the still limited number of studies investigating the on-site success of SCs, we do not expect that inferential statistics generate reliable results. Rather, the statistical approach serves us to explore the robustness of the patterns we identified in the descriptive analysis.

In order to ensure a high-quality analysis, throughout the coding process, each article was coded separately by two coders. When coders did not agree on either the occurrence of a category in an article or the evaluative statement's value (on a 1–3 scale as previously discussed), the coders convened to discuss their decisions until they came to a shared decision. In doing so, we were able to agree on all coding decisions.

4. Results

4.1. Evidence base

Our review shows how frequently different research approaches were used in our sample of studies (see Fig. 3b,c,d). With 68 cases, quantitative (quasi-)experimental approaches were the most common research approach, closely followed by qualitative (n = 67), and quantitative non-experimental approaches (n = 55). The differentiation of the literature according to data sources shows a relatively even distribution between the use of documents (n = 88), direct observations (n = 77), surveys (n = 68) and in-depth interviews (n = 75). Significantly fewer articles (n = 38) used panel data to analyze the on-site success of SCs over time.

Since the early 2000s, the number of annual publications that have investigated the success of SCs in improving the sustainability of certified producers has been has been rising (see Fig. 4). For example, while our literature base counts less than 5 publications in 2005, it contains 24 publications in 2019 and 19 publications in 2020.

Strikingly, the distribution of publications is highly uneven across individual SCs (see Fig. 1). The majority of publications investigated the FSC (n = 61) and FT (n = 38). Also, the MSC, RFA and FT.org feature relatively prominently in our literature base with around 20 publications followed by RSPO and UTZ with around 15 publications. However, for nine of the 18 SCs included in our sample, our literature base comprises less than five research papers each. For one SC, Cotton Made in Africa (CMA), our search yielded not one publication suited to our parameters. As for outcome variables, the reviewed articles have addressed

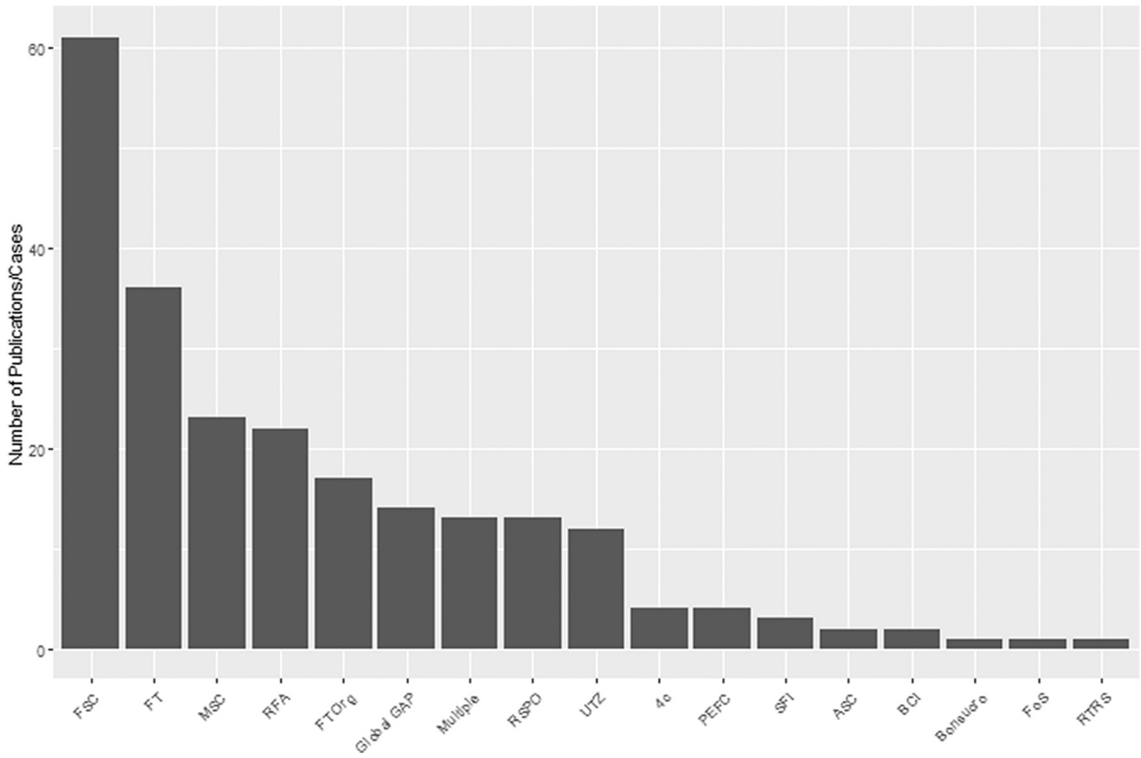


Fig. 1. Distribution of studies included in the review over individual SCs included in Table 1.

environmental (n = 154), social (n = 132), and economic (n = 150) outcome variables at relatively similar frequencies.

Regarding context, among the three sub-sectors, SCs operating in farm-agriculture (n = 92) have been studied the most, followed by SCs operating in the forestry sector (n = 60). Considerably fewer studies have focused on SCs that operate in the seafood sector (n = 25) (see Fig. 6c). Timber (n = 60) is the single most studied commodity followed by coffee (n = 32) and seafood (n = 25) (Supplementary material 2). There are considerably fewer case studies that focus on other commodities, such as cacao, palm oil or tea. As for the distinction between smallholder and non-smallholder production sites (see Fig. 6d), we see a relatively even distribution of studies focusing on smallholder (n = 83) and non-smallholder (n = 76) production sites.

Fig. 2 shows the distribution of studied cases over countries and regions. On the level of individual countries, most cases address production sites in Kenya and Indonesia. On a regional level, Latin America has the highest number of cases. Together with North America, Latin America is also the region with largest geographical coverage. In term of cases, Latin America is followed by Africa and Asia. However, country coverage is lower here than in Latin America. Also, coverage in Europe is

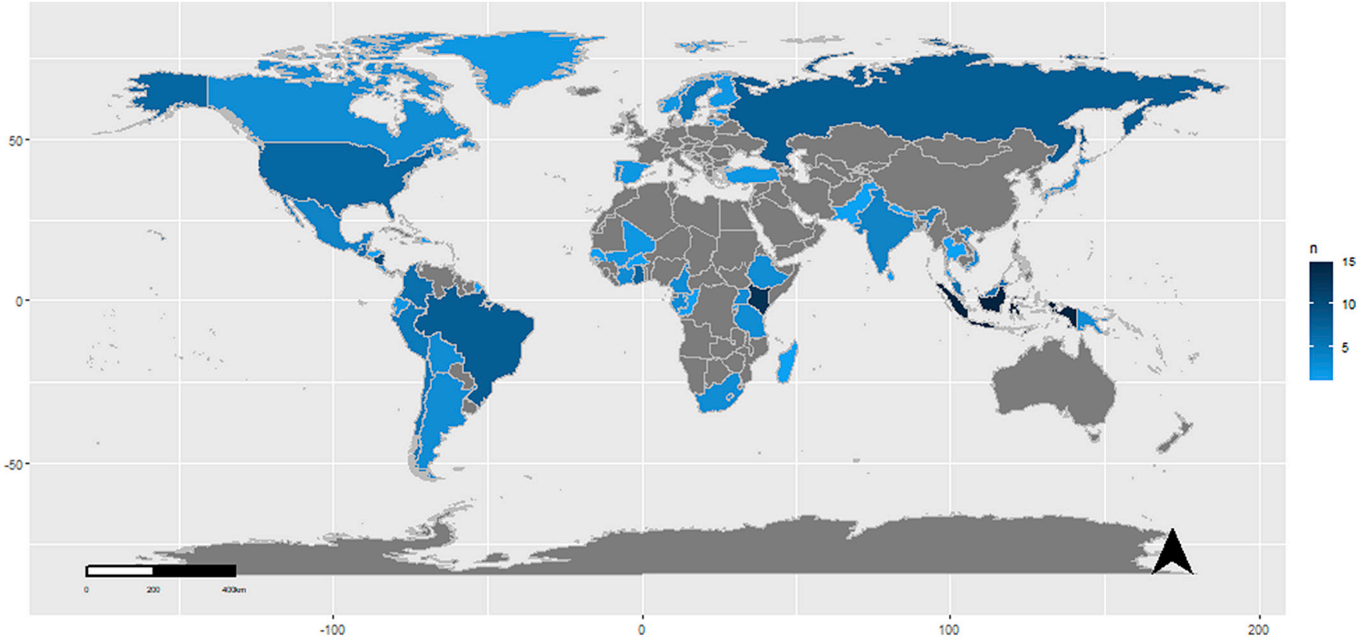


Fig. 2. Distribution of studies across countries. The color indicates the number of studies, no studies were recorded for gray areas.

lower, with cases focused on the Scandinavian countries, followed by Portugal, Spain and Turkey.

Finally, our review shows a relatively even distribution of cases between intermediate ($n = 228$) and endpoint variables ($n = 207$) as well as between outcome variable with low ($n = 45$), medium ($n = 43$) and high ($n = 58$) implementation costs (see Fig. 8a and b).

4.2. *Evidence gaps*

Overall, the evidence base has grown steadily over the past two decades, and there is no clear sign that interest in SC-related issues would have peaked. The literature adopts a broad perspective by using a variety of different research approaches and data sources to study diverse outcome variables. However, parts of the evidence base remain fragmented. The most striking research gaps concern the uneven distribution of case studies across different SCs, where we find that the literature is biased towards assessments of the FSC and FT, while other important SCs received considerably less attention. Profound evidence gaps exist due to missing studies for almost half of the SCs in our sample including major schemes such as 4C, the leading (in terms of volumes) certification scheme in the global coffee market. Also, the farming-agriculture and forestry sectors received significantly more attention than the seafood sector, while on the level of individual commodities we find a bias towards coffee and timber. Last but not least, country coverage remains limited in most world regions.

After analyzing our evidence base, we now proceed to answer the research questions posed in the introduction.

4.3. *On-site success of SCs: An overview*

How successful are the major SCs in improving the on-site sustainability performance of certified production sites? Fig. 3a shows the distribution of favorable, mixed, and skeptical evaluations over all articles in our literature base. In sum, the results point to limited success of SCs. The picture is dominated by skeptical evaluative statements, which comprised 38.5% of all studies, followed by mixed evaluations (36.8%). Only 24.7% of studies in our literature base conclude with a favorable evaluation.

Among the different research designs that are used to evaluate the success of SCs at producer level, quasi-(experimental) studies are widely considered to present the most rigorous approach. As shown by Fig. 3b, the majority of studies (38%) that use a quasi-experimental approach close with a skeptical evaluation followed by mixed (37%) and favorable (25%) evaluations. Interestingly, qualitative approaches show highly similar results. Also, these studies are dominated by skeptical evaluations. The picture, however, differs for quantitative observational studies which show a larger, albeit, statistically insignificant, share of favorable observations. More rigorous quantitative approaches that control for selection bias, thus, tend to evaluate the success of SCs more skeptically than less rigorous quantitative approaches. However, we found that these differences are not statistically significant (see Supplementary material 3 and 4).

Further, Fig. 3c shows the distribution of favorable, mixed, and skeptical evaluations across case studies that use different data sources (i.e. surveys, direct observation, in-depth interviews). The picture is clearly dominated by skeptical evaluations for cases that are based on document analysis (39%) in-depth interviews (45%) and panel data (45%). For cases that are based on direct observations, mixed (36%) and skeptical distributions (35%) show similar results. Only survey-based case studies show a different distribution, with mixed results dominating the picture (43%), followed by favorable (31%) and skeptical evaluations (26%). Comparing the distribution of evaluations across cases using Chi^2 Tests shows no statistically significant differences (see Supplementary material 3). However, the binominal regression model shows that articles based on surveys are significantly more likely to conclude with a favorable conclusion while articles based on panel data are significantly more likely to conclude with a skeptical evaluation (see Supplementary material 4). Also, articles based on interviews are significantly more likely to end with a skeptical conclusion (see Supplementary material 4). Articles using direct observations, on the other hand, are more likely to reach a favorable conclusion.

However, most strikingly, we found a largely consistent distribution of evaluations across the case studies within the eight different subsamples related to research design. In all but two sub-samples the number of cases that conclude with a skeptical evaluation is higher than the number of cases that end with a favorable evaluation. Nevertheless,

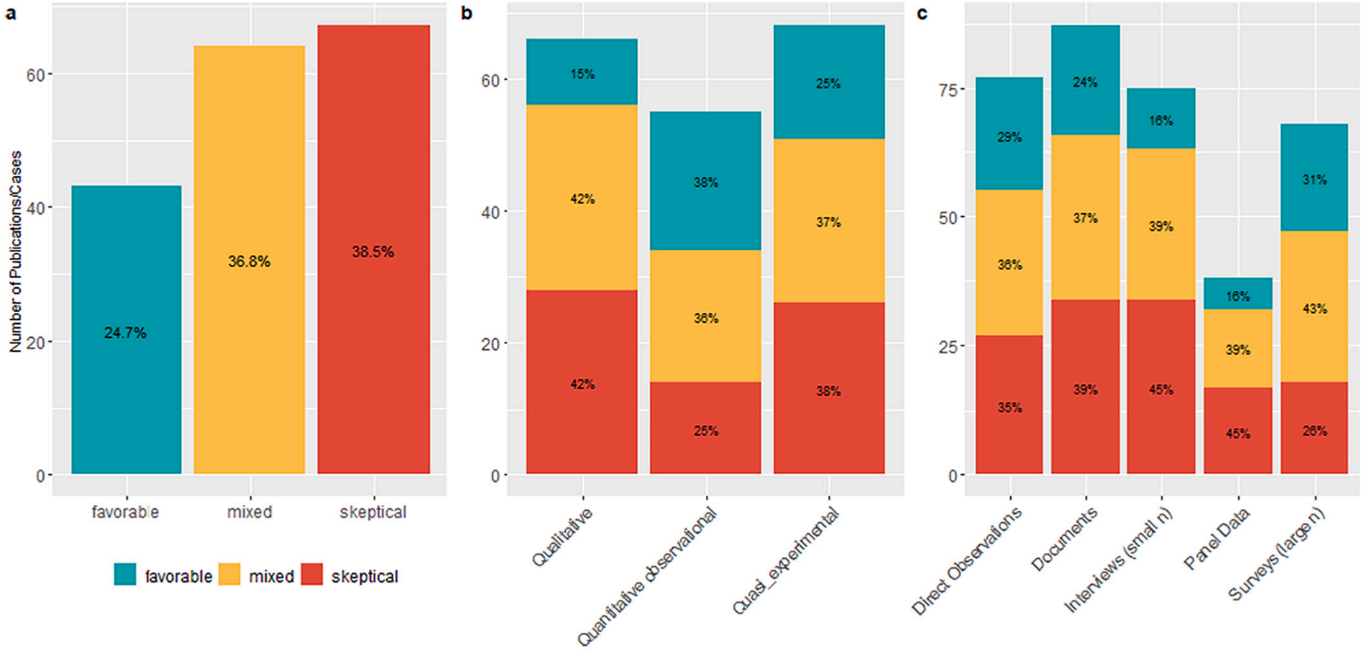


Fig. 3. Distribution of favorable, mixed and skeptical evaluations for all articles (a), across different methodological approaches (b), summarized methodological categories (c) and data sources (d).

the findings for both the categories of quantitative observational research approaches and surveys suggest that cases using these methods may be biased towards favorable conclusions which, in turn, suggests that the use of less rigorous research designs tends to lead to more favorable evaluations. In addition, we examined possible effects of the type of journals in which the articles were published on the overall ratings but found no significant differences (see Supplementary material 8 and Supplementary material 3).

4.4. *Success of SCs over time (dimension of time)*

Given the rapid growth in the number of SCs, the sectors they cover and their ever-expanding geographical coverage, we asked whether the evaluations of SCs have changed over time, and if so, to what extent? The distribution of evaluative statements over time illustrates with a few exceptions that skeptical evaluations increasingly dominate over favorable evaluations. Interestingly, we find a period of five years (2011–2015) during which the distribution of favorable, mixed and skeptical evaluations was almost equal (Fig. 4). However, as the number of annual publications increased in the years that followed (2016–2020), so did the proportion of skeptical evaluations. For the period of 2016 to 2020, our analysis shows the largest share was comprised of skeptical evaluations (41%) followed by a 38% share of mixed results and only 21% for favorable assessments. Overall, our review therefore indicates a relative increase of skeptical evaluations over time.

4.5. *Success of SCs differentiated for the categories of individual SCs, different types of SCs, different outcome variables and different contexts (dimension of scope)*

How does the evaluation of the success of SCs vary? First, we break down the distribution of evaluations to the level of individual SCs (Fig. 5). As a caveat, it needs to be noted that the number of cases is far too small to produce reliable results for many of the included SCs (below five cases). The most reliable results concern FSC and FT, since they stand out as the two most studied individual SCs. With 39% (FSC) and 44% (FT), the results for these two SCs are dominated by skeptical evaluations. It is important to note that Rainforest Alliance (RFA) and Global GAP are evaluated more positively, although this remains statistically insignificant (see Supplementary material 3). Cases that study the sustainability performance at RFA-certified producers show a share of 41% positive evaluations. For Global GAP-certified production sites this value even increases to 50%. Overall, differences between SCs were not significant (see Supplementary material 3).

Second, in Table 4 we grouped the individual SCs into one group of NGO/civil society-driven SCs, and another group for corporate/business association-driven SCs. NGO/civil society-driven SCs show a higher share (25%) of favorable evaluations than corporate/business association-driven schemes (16%). Nevertheless, the majority of evaluations for both types of SCs is mixed (Fig. 6a). In sum, we find no statistically significant differences in evaluation outcomes between the two types of SCs. (see Supplementary material 3).

We distinguish the articles further according to the area of sustainability they address. We assign all articles that study outcomes related to prices, producer income, producer well-being, poverty reduction or productivity to the area of economic sustainability. All articles that study the improvement of living or working conditions for workers or other disempowered or marginalized groups such as women, children or indigenous people are assigned to the area of social sustainability. Finally, we assign all articles all that focus on the effects of SCs on reducing the ecological footprint of primary production to the area of environmental sustainability. We did not break down the analysis further to individual outcome variables within the areas of economic, social and environmental sustainability, since for individual outcomes the number of pertinent studies is mostly too small to allow for aggregate conclusions.

Fig. 6b suggest that SCs are slightly, statistically not significant, (see Supplementary material 3) more successful in addressing issues concerning social sustainability, when compared to economic and environmental issues. For case studies that assessed social outcome variables, a majority of 40% conclude with favorable evaluations, followed by skeptical (34%) and mixed (26%) evaluations. However, for both economic and environmental outcomes this distribution is different. For these sustainability pillars, skeptical evaluations (40%, 44%) clearly dominate the picture, followed by mixed evaluations (30%, 30%). Favorable evaluations (31%, 27%) once more fall behind. Below, we further differentiate this analysis and show results separately for intermediate and endpoint outcomes (see Fig. 8a). It becomes clear that the higher proportion of favorable evaluations in the dimension of social sustainability only applies to the type of intermediate outcomes. For endpoint outcomes - which are more significant for the overall evaluation - the majority of evaluations are also negative in the dimension of social sustainability.

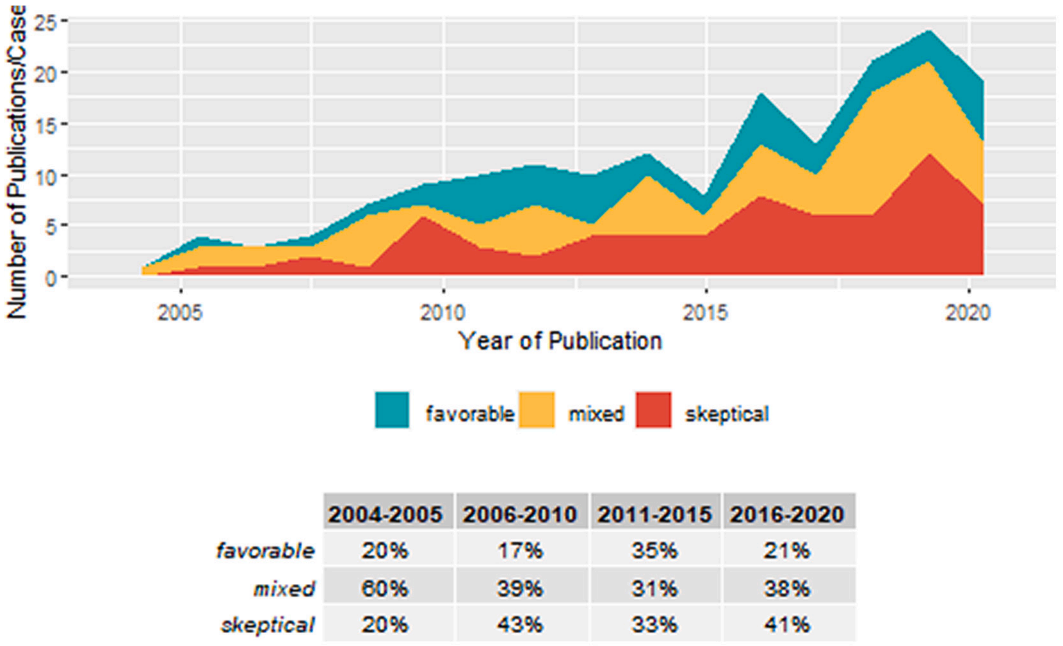
In order to assess potential tradeoffs and synergies between the three sustainability dimensions we evaluated tendencies of studies assessing more than one sustainability dimension. Due to the relatively low number of studies that assess sustainability outcomes across sustainability dimensions this analysis is limited and generated no statistically significant results. The results that emerge from our analysis, however, suggest that synergies between different sustainability dimensions are more likely than trade-offs. If a case ends with a favorable conclusion in one dimension, it is also - at least slightly - more likely to end with a favorable conclusion in a second dimension. This trend is most evident in the analysis across the areas of economic and environmental sustainability (Supplementary material 7).

Contextual factors affected the analysis much less than we anticipated. Fig. 6c clearly illustrates that the sampled literature evaluates the success of SCs largely similarly, independent of the subsector (forestry, farm-agriculture, fisheries/aquaculture) to which a particular case refers. Similarly, we do not find any statistically significant differences between individual business sectors (Supplementary material 2, 3 and 4). Further Fig. 6d shows that the distribution of evaluations does not vary significantly (see Supplementary material 3 and 4) between cases that focus on smallholder production sites as opposed to cases that focus on non-smallholder production sites. Independent of the type of production site a case focuses on, the picture is dominated by skeptical evaluations (42%, 43%), followed by mixed (30%, 33%) and favorable (28%, 24%) evaluations.

Moreover, national conditions ranked according to indicators for land use (share of agricultural area, agricultural value added, Fig. 7a; agricultural area, Supplementary material 5 - Fig. 5a), human development (Human Development Index, Fig. 7b), governance (fragile states index, Fig. 7c), environmental performance (environmental performance index, Fig. 7d), international trade (share of agricultural commodities, Supplementary material 5 - Fig. 5b), and consumption patterns (biocapacity and ecological footprint, Supplementary material 5 – Fig. 5c and d), did not have a significant impact on how the literature evaluates the success of SCs. Boxplots show a similar distribution of favorable, mixed and skeptical evaluations across all tested country indicators without any significant differences, as did Paired Wilcox Tests with Bonferroni correction (Fig. 7, Supplementary material 6).

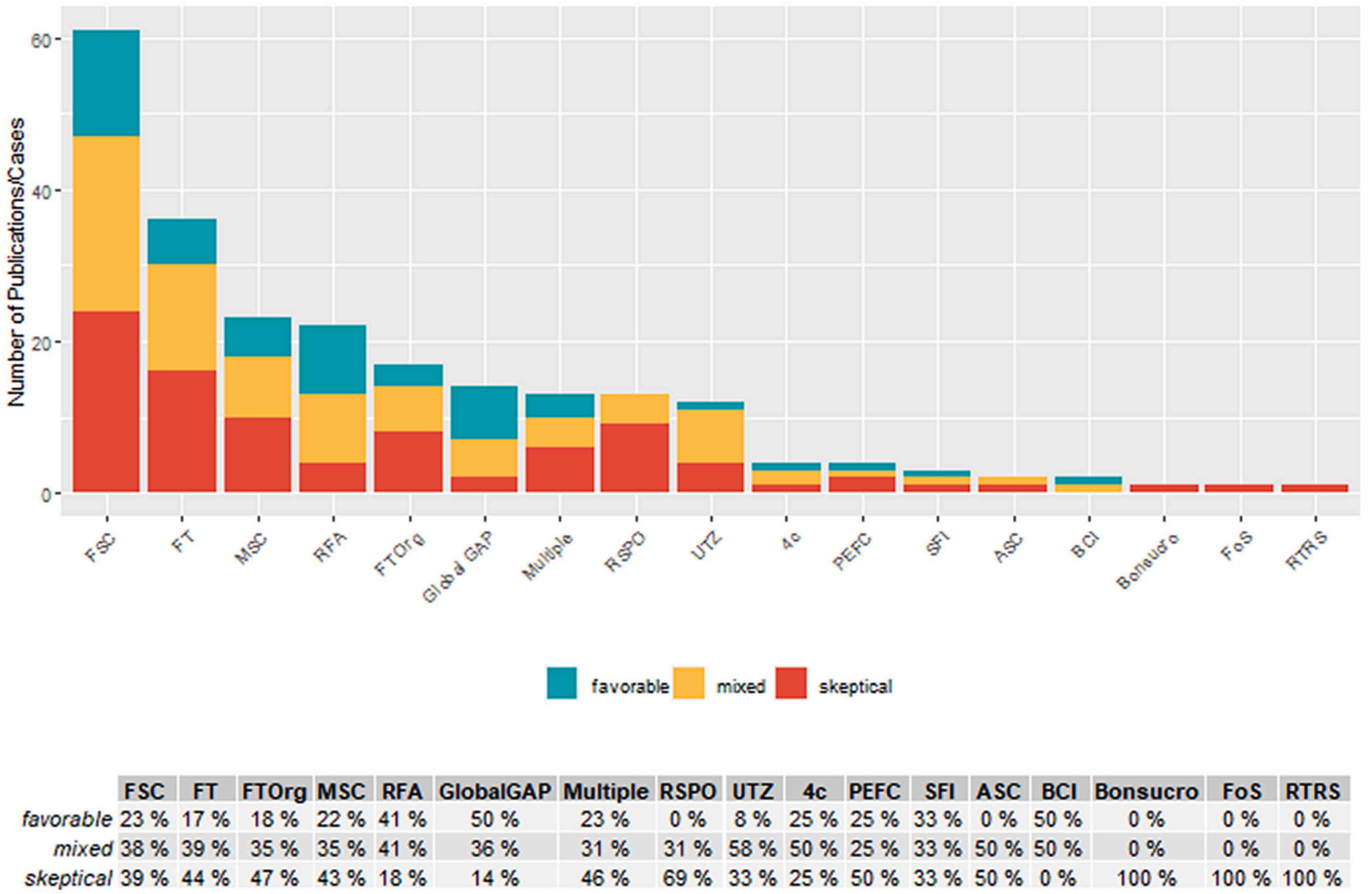
4.6. *The success of SCs differentiated for the categories of intermediate and endpoint variables and for the categories of low, medium and high implementation costs (dimension of depth)*

Finally, we assess the success of SCs along the dimension of depth. How profoundly do SCs improve the sustainability performance of certified producers? Notably, the evaluations change considerably between intermediate and endpoint outcome variables. As explained above, while intermediate outcome variables are means to an end (e.g.,



	2004-2005	2006-2010	2011-2015	2016-2020
favorable	20%	17%	35%	21%
mixed	60%	39%	31%	38%
skeptical	20%	43%	33%	41%

Fig. 4. Distribution of favorable, mixed and skeptical evaluations of all studies across time with summarized distribution values lumped for 5-year periods.



	FSC	FT	FTOrg	MSC	RFA	GlobalGAP	Multiple	RSPO	UTZ	4c	PEFC	SFI	ASC	BCI	Bonsucro	FoS	RTRS
favorable	23 %	17 %	18 %	22 %	41 %	50 %	23 %	0 %	8 %	25 %	25 %	33 %	0 %	50 %	0 %	0 %	0 %
mixed	38 %	39 %	35 %	35 %	41 %	36 %	31 %	31 %	58 %	50 %	25 %	33 %	50 %	50 %	0 %	0 %	0 %
skeptical	39 %	44 %	47 %	43 %	18 %	14 %	46 %	69 %	33 %	25 %	50 %	33 %	50 %	0 %	100 %	100 %	100 %

Fig. 5. Distribution and percentage values of favorable, mixed and skeptical evaluations differentiated for all individual SCs included in Table 1.

price premiums are a means by which smallholders can be lifted out of poverty), endpoint variables reflect ultimate sustainability goals linked to sustainability goals such as poverty reduction, climate change mitigation, or reduction of species losses. Fig. 8a shows that cases which analyze intermediate outcome variables evaluate the success of SCs considerably more favorably than cases that focus on endpoint outcomes. The share of favorable evaluations lies at 41% for evaluations of intermediate outcome variables, whereas for endpoint outcomes, only

Table 4
Types of SCs.

NGO/Civil Society-Driven	Corporate/Business Association-Driven
FoS, FT, FT.org, RFA, FSC, MSC, RTRS, ASC	SFI, 4C, CMA, RSPO, BCI, UTZ, GAA, GlobalGAP, PEFC, Bonsucro

22% of evaluations turned out favorably. Strikingly, these differences are found statistically significant in both the Chi^2 Tests (see Supplementary material 3) and the binominal regression model (see Supplementary material 4).

This general difference becomes especially visible in the area of economic sustainability. Here, for cases that study intermediate outcome variables, a significant majority of 42% conclude with a favorable evaluation, in comparison to only 16% of the cases that study endpoint outcome variables. Instead, the picture for endpoint economic variables is clearly dominated by negative evaluations (46%). In the area of social sustainability (Fig. 7a), the gap in assessments between cases that focus on intermediate and endpoint outcome variables is similarly striking . Chi^2 Tests show that the different distribution of favorable, mixed and skeptical evaluation between economic intermediate and economic endpoint outcome variables is statistically significant (see Supplementary material 3).

In terms of environmental sustainability, the picture painted by the literature becomes even more discouraging. The distribution of these evaluations is dominated by skeptical evaluations for both types. In addition, also in this area the share of skeptical evaluations is significantly higher (10 points) for cases that focus on endpoint outcomes if

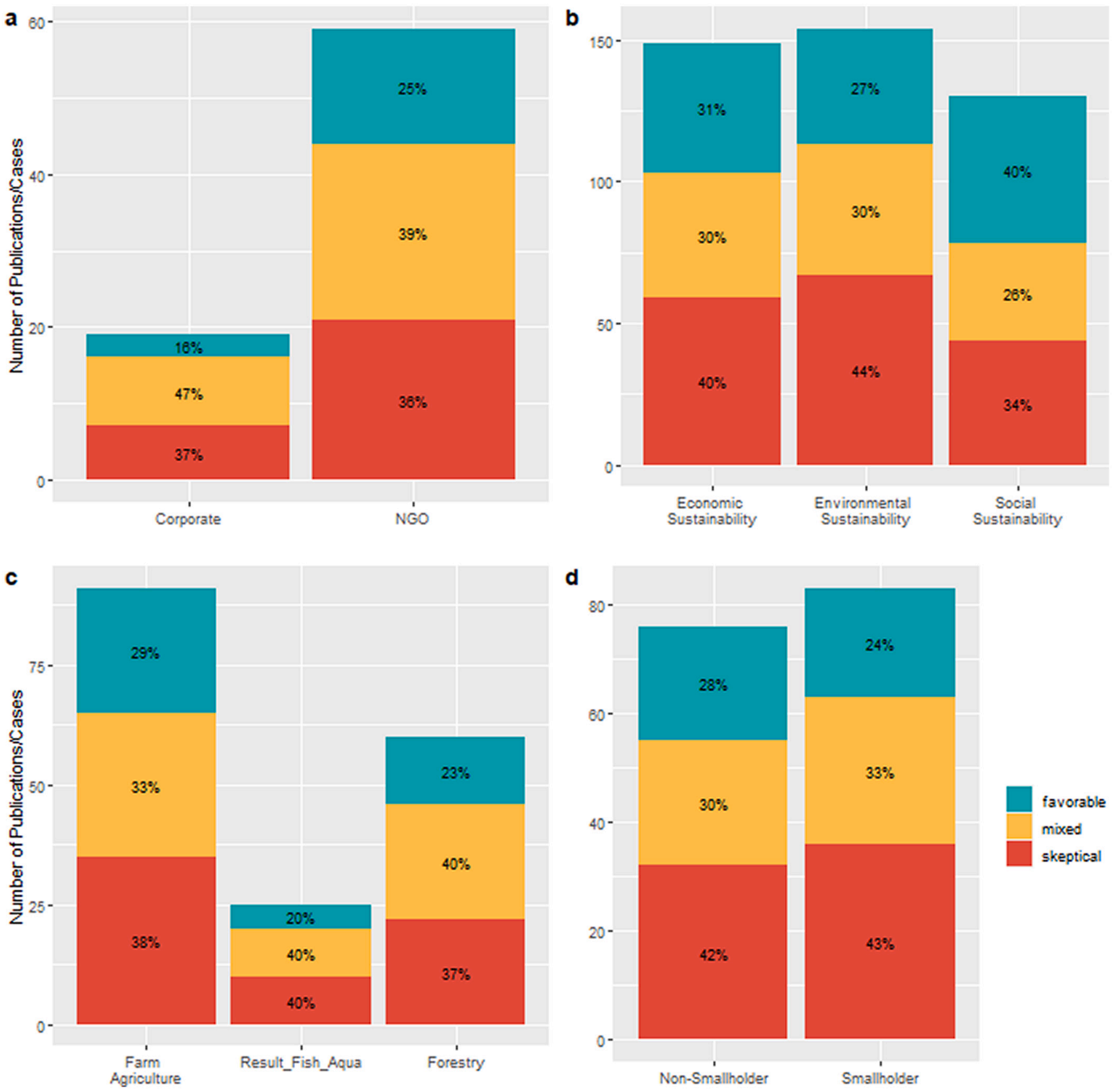
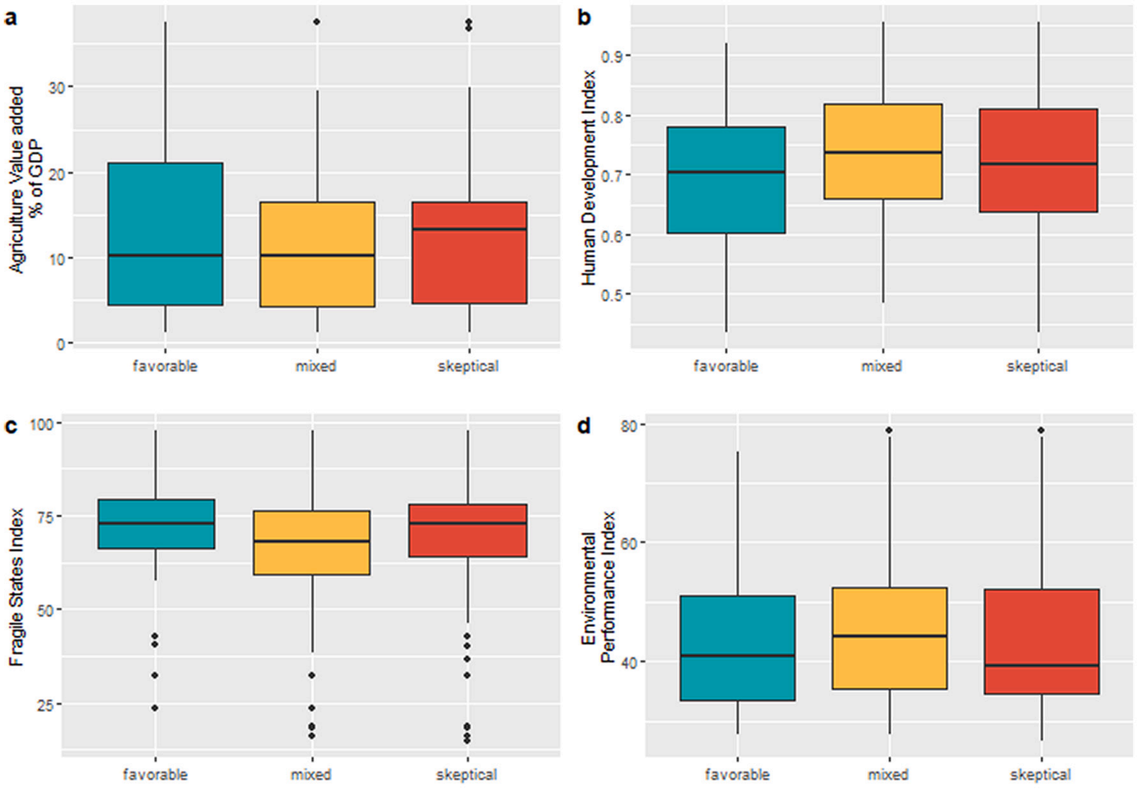


Fig. 6. Distribution of favorable, mixed and skeptical evaluations differentiated for types of SCs (a), outcomes variables in the three different areas of sustainability (b), type of economic sector (c) and development status of certified production site (d).



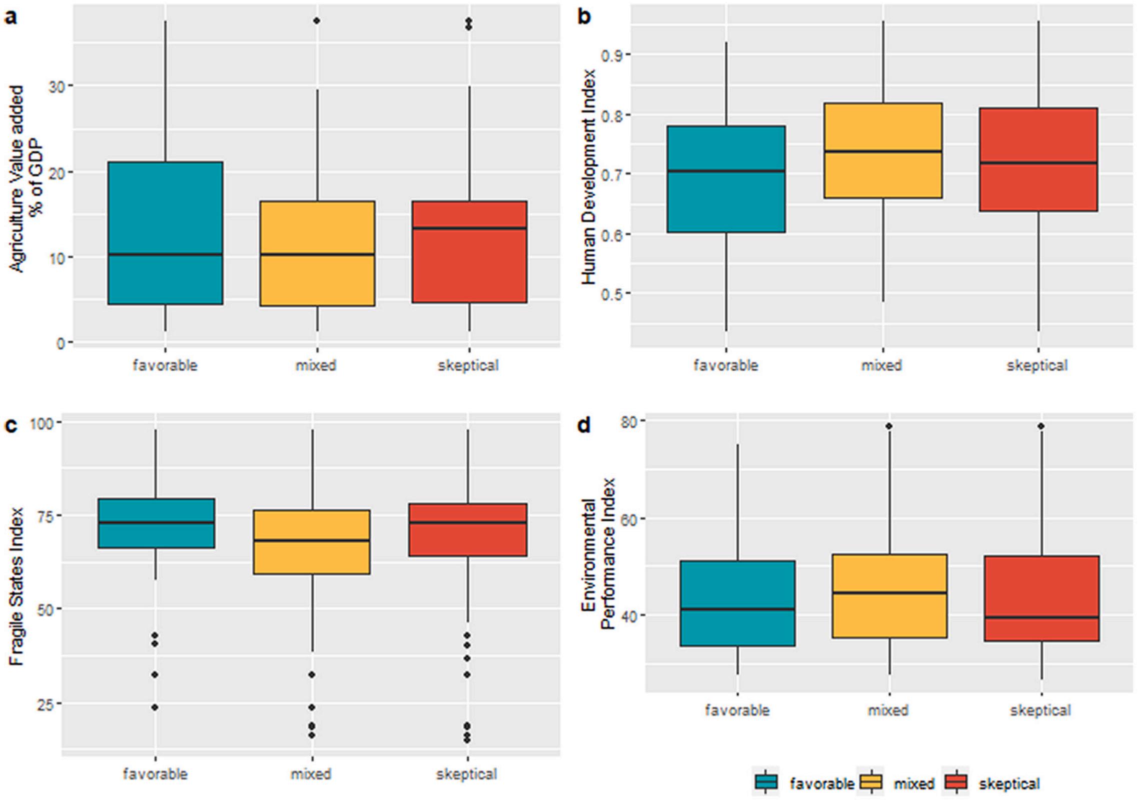


Fig. 7. Variation of favorable, mixed and skeptical evaluations for national conditions: Agricultural Value Added as percentage of national GDP (a) The World Bank (2020) Agriculture, forestry, and fishing, value added (% of GDP) World Development Indicators. The World Bank Group, https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS, downloaded 14.10.2020, Human Development Index 2019 (b) United Nations Development Program (2019) Human Development Index 2019, http://hdr.undp.org/en/data, downloaded 22.1.2021, Fragile States Index (c) Fund for Peace (2020) FRAGILE STATES INDEX ANNUAL REPORT 2020, https://fragilestatesindex.org, downloaded 4.10.2020 and Environmental Performance Index (d) Wendling, Z. A., Emerson, J. W., de Sherbinin, A., Esty, D. C., et al. (2020). 2020 Environmental Performance Index. New Haven, CT: Yale Center for Environmental Law & Policy. epi.yale.edu, downloaded 4.10.2020.

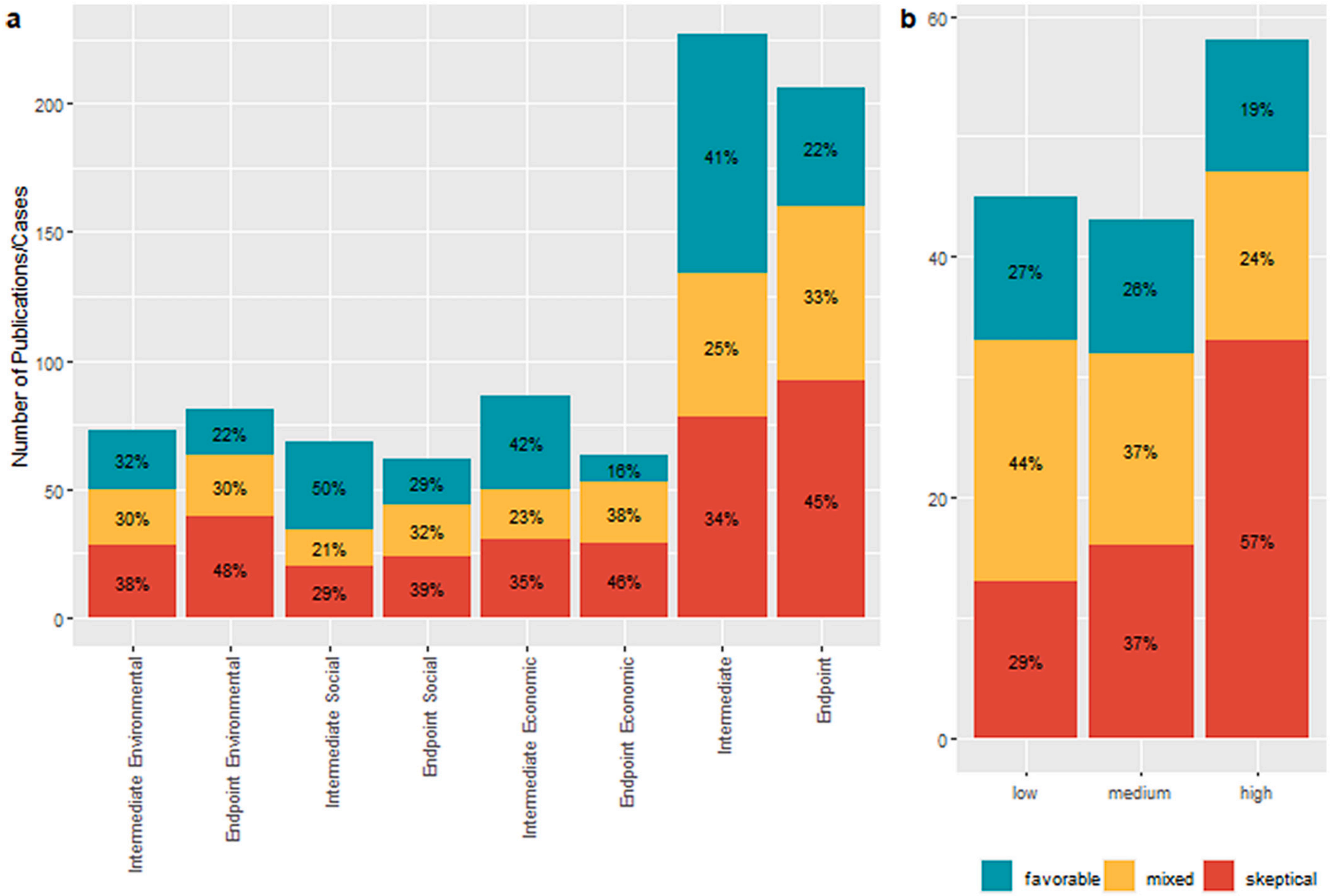


Fig. 8. Distribution of favorable, mixed and skeptical evaluations differentiated for the categories of intermediate and endpoint variables overall and within different areas of sustainability (a), distribution of positive, mixed and negative evaluative statements differentiated for the categories of low, medium and high implementation costs(b).

compared to cases that focus on intermediate outcome variables. Again, Chi^2 Tests show that these differences are statistically significant (see Supplementary material 3). Overall, the less profound type of intermediate outcome evaluations are significantly more favorable than evaluations of the more profound endpoint outcome across all sustainability dimensions.

As for the distinction between outcome variables with low, middle and high implementation costs, Fig. 8b shows that the share of skeptical evaluations increases the higher the implementation costs, while simultaneously, the share of favorable assessments decreases. Outcome variables with low implementation costs exhibit a share of 29% skeptical evaluations, 27% favorable evaluations and 44% mixed evaluations. For cases that focus on outcome variables with medium implementation costs, the share of skeptical evaluations lies at 37%, while the share of favorable evaluations is marginally lower at 26%.

However, for cases that analyze the success of SCs on outcome variables with high implementation costs, these values change dramatically. For such cases, the share of skeptical evaluations increases to a share of staggering 57%, while at the same time, the share of favorable evaluations decreases to 19%. These differences are statistically significant according to the Chi^2 Test (Supplementary material 3). Moreover, the binominal regressions model confirms that articles that investigate outcome variables with high implementation costs are significantly more likely to reach a skeptical conclusion (see Supplementary material 4). Overall, these analyses paint a highly negative picture for the in-depth dimension of SC's success.

5. Discussion

It is of critical importance for scholars, policymakers and industry professionals to understand whether SCs are effective enough to drive or bolster the sustainability transformations so urgently needed to safeguard global bio-based primary production in the present and coming decades. This study aims at improving the underlying knowledge base for the debate on what role SCs may play in transformation governance. While some experts continue to view the SC model as a successful governance vehicle, others have criticized them as mere greenwashing instruments, incapable of bringing about profound change. Indeed, support for both these positions can selectively be found in the relevant literature. The ground-level studies that have investigated the success of SCs at certified production sites have reached favorable, mixed, and skeptical evaluations. Critically, however, in our meta study skeptical evaluations prevail. Based on a broader evidence-base than many recent quantitative meta-studies, we thus draw a largely discouraging picture about the success of SCs at producer level. All in all, this pattern appears very similar across the primary production sub-sectors of farm-agriculture, forestry, and fisheries/aquaculture in which SC plays a major role. Our binominal regression model (Supplementary material 4) shows that case studies relating to Europe have a higher probability of favorably evaluating the performance of SCs. This can be interpreted as an indication that SCs work particularly well where stricter environmental and social legislation exists anyway.

One of the most striking findings of our review is the clear differentiation in outcomes between cases that focus on intermediate vs. endpoint outcomes, as well as those that focus on outcomes with high, middle and low implementation costs. The pattern that has emerged from the data demonstrates that evaluations focusing on less rigorous outcome variables, or outcomes with lower implementation costs, comprise a disproportionate share of favorable outcome evaluations. Recent, quantitative meta-studies by Garrett et al. (2021), Meemken (2020) and Oya et al. (2018) have observed similar trends. Meemken (2020) observed that certified farmers on average received significantly higher prices (intermediate outcome) for their produced commodities, but that this increase in prices had only small effects on household incomes (endpoint outcome) and almost no or even negative effects on poverty reduction (endpoint outcome). Also, Oya et al. (2018) found a comparatively high impact of SCs on prices (intermediate outcome) but smaller impacts of SCs on household incomes (endpoint outcome) and wealth (endpoint outcome) and even negative effects on wages – an outcome variable, whose implementation according to our coding scheme requires high implementation costs. From an environmental perspective, Garrett et al. (2021) and Börner et al. (2020) looked into effects of SCs on deforestation and reductions in the use of fire. Both studies found that SCs only have a comparatively small impact on reducing deforestation rates and Garrett et al. (2021) also found no measurable reduction in the use of fire. Overall, recent meta-studies, therefore, seem to converge on the finding that SCs are particularly limited in producing profound and additional sustainability gains.

An important puzzle that emerges from our research is why a successful implementation of intermediate outcomes does not materialize to the same extent in a successful implementation of endpoint outcomes. One possible explanation for this may be goal conflicts between different economic, environmental and social sustainability outcomes so that success in one area of sustainability compromises the achievement of sustainability goals in another area (Vanderhaegen et al., 2018; Marx et al., 2021). For example, the price premiums certified producers receive for their standard compliant products may be too small to reimburse them for increased regulatory costs (Dietz et al., 2020). Further, sustainability governance is a complex task. SCs cannot control all social interactions that ultimately have an effect on sustainability outcomes (Wijen, 2014). Bennett (2022), for example, argues that certification may lead to a successful implementation of grievance systems in certified farms (intermediate outcomes). However, in practice, these systems are almost never used, which is why their impact on improved human rights practices remains largely limited (endpoint outcome). In a similar vein, Grabs (2020) found that certification holders largely follow the certification requirement to provide protection gear to farm workers engaged in spraying activities (intermediate outcomes). But it was equally evident that in many cases farm workers did not make use of the protective clothing, or at least not properly, in their daily routines in order to improve their health conditions (endpoint outcome). Finally, scholars have pointed towards negative spill-over effects to explain the missing success of SCs. In such cases, unsustainable practices are eliminated at the certified production unit but only to be shifted to another uncertified production site, leading to largely unchanged overall sustainability outcomes (Bastos et al., 2019).

Important differences exist between our qualitative approach and established quantitative meta-study approaches that may help explain the relatively skeptical picture that emerges from our study. While scholars in original studies who find statistically negative or insignificant effects usually also evaluate the overall success of SCs negatively, it is not uncommon that scholars who measure significant positive effects, nevertheless evaluate the success of SCs critically after they contextualized and interpreted their findings. This may, for example, be the case when the measured effects are relatively small. To illustrate this point, consider the following quotation taken from Weber (2011, p 677) summarizing his study results: "FT-organic growers received an average premium of 12.8 cents per pound, yielding a gross income gain of 5% of total household income or about 26 dollars per household member. The gain is net of the costs of cooperative participation but not of other costs incurred to become certified and suggests that price premiums alone have a limited potential to increase household returns from growing coffee. More broadly speaking, the finding raises questions about the persistence of substantial price premiums associated with social or environmental labeling initiatives."

Further reasons for a discrepancy between statistically measured effects and final evaluation may be of a contextual nature. Consider the following example by Carlson et al. (2018, 121): "While forest loss and fire continued after RSPO certification, certified palm oil was associated with reduced deforestation. Certification lowered deforestation by 33% from a counterfactual of 9.8 to 6.6% y−1. Nevertheless, most plantations contained little residual forest when they received certification. As a result, by 2015, certified areas held less than 1% of forests remaining within Indonesian oil palm plantations. […] Broader adoption of certification in forested regions, strict requirements to avoid all peat, and routine monitoring of clearly defined forest cover loss in certified and RSPO member-held plantations appear necessary if the RSPO is to yield conservation and climate benefits from reductions in tropical deforestation."

In a quantitative vote-counting design with an exclusive focus on numbers, the findings of the two cited studies would push the average effect sizes in a positive direction. However, within our qualitative text-based study-design both studies would increase the relative share of skeptical or mixed evaluations. In other words: Depending on the research lens, and potentially also ideology bias among authors, the same study may lead to different insights depending on the meta-study approach. While quantitative meta-studies are well equipped to aggregate effect sizes across studies, they lack the ability to synthesize the evaluative conclusions that original studies draw based on their empirical insights. Quantitative meta studies may therefore face the risk of misinterpreting the positive effects of SCs in cases where statistically measurable positive impacts do not translate into tangible sustainability improvements. At least partly, these differences in study-design may explain why our study is more skeptical about the success of SCs than related quantitative studies.

What factors might explain the limited success of SCs? In the SC literature, it is often expected that certain case-specific attributes — such as SC type, the development level of the study site (Sellare et al., 2020), or country context (Eberlein et al., 2014) —have a significant impact on their success. The patterns that we found in this review, however, show a largely similar domination of skeptical evaluations across these different categories. In this respect, our findings are largely consistent with the findings by Meemken (2020) who also found that contextual factors typically accounted for by the pertinent literature do not explain much of the variation in SCs' success.

Our results, therefore, tend to support explanations that criticize the mechanism of SCs as such. Many studies have acknowledged that the success of SCs on the ground hinges on the rules they promote – both the substantial standards and the rules that determine the certification and auditing process. While some authors have portrayed SCs as inclusive multi-stakeholder initiatives with quasi-democratic structures (Meidinger, 2011; Mena and Palazzo, 2012) that use experimental governance techniques (Overdevest and Zeitlin, 2014; Rasche, 2012) to create adaptable rule systems, others have developed a much more skeptical view. Many critical views regard the coexistence of different SCs as a major problem since it leads to competition and may cause a regulatory race to the bottom. (Fransen, 2012; Egels-Zandén and Wahlqvist, 2007; Bloomfield, 2012; Dietz et al., 2018). Dietz et al. (2018) have drawn on the example of the global coffee industry to show that less strict SCs have grown significantly faster than SCs with stricter standard systems and called this a "relative race to the bottom". Further, scholars have argued that the rule-setting processes of certification organizations are only inclusive on paper while the real process happens backstage where the important political decisions are taken (Hatanaka et al., 2012). In a recent article, Bartley (2022) stated that standard development within current SC systems has become largely dominated by corporate interests. In an earlier article, Reinecke et al. (2012) argued that SCs only superficially pretend to confront global companies as opposing actors. In reality, both SCs and global companies are driven by the same overriding logic of developing successful business models that will prevail in the market. This may also explain our finding that for both types of SCs (NGO/Civil Society-Driven and Corporate/Business Association-Driven) skeptical evaluation prevail (see Fig. 6).

Indeed, market success of SCs (also often called mainstreaming) has been presented as an ambiguous concept. On the one hand, it drives standard diffusion and may, thus, change the practices of powerful actors (Pattberg, 2005). However, on the other hand, it may also undermine the effectiveness of SCs as large actors capture the rule-setting dynamics within SCs systems. (Llach et al., 2015; Ponte, 2014; Ponte, 2012a). For example, Ponte (2012a) has argued that increasing the adoption of MSC fishery certifications does not necessarily imply that fish-stocks are recovering in practice. Building on these insights, Grabs (2020) and Dietz and Grabs (2021) have argued that under these conditions SCs face a difficult choice. Either they weaken their standards to increase adoption or they will be pushed out of market. The financial dependence of SCs on standard adoption may create incentives to weaken standards in order to attract clients (Fortin and Richardson, 2013; O'Rourke, 2006; Ponte, 2012b). Overall, the patterns that emerged from our review tend to support these views. Especially in recent years, the distribution of evaluations has trended towards skeptical evaluations (see Fig. 4). This may reflect a dynamic in which SC claims and real effects on the ground increasingly decouple from each other (Barrientos and Smith, 2007; Dietz et al., 2019; Wijen, 2014).

6. Conclusion

Primary production systems, essential for human survival, are a major factor of the earth system exceeding planetary boundaries. Agriculture expansion is a major driver of deforestation (Pendrill et al., 2019). In the tropics, rainforests, savanna and other ecosystems who are home to the most biodiverse habitats on earth are replaced through new agricultural land. Agriculture also accounts for more than 70% of the global freshwater withdrawals.[1] Fish stocks are decreasing, while the excessive use of agro-chemicals leads to soil, air and water pollution (Sachs, 2015). Moreover, most of the global poor live in rural areas suffering from low productivity and poor working conditions including child work and slavery (Sachs, 2015). The Covid-19 pandemic and the current geopolitical crisis make this situation even worse by interrupting global commodity chains and undermining global food production. These are but only a few examples illustrating the need for the development of an effective governance framework that reconciles economic goals with ecological and social aspects in existing global primary production systems.

SCs are now routinely discussed as one potential governance tool to address these issues. However, if we take on a problem-solving perspective and benchmark the effects of SCs on the ground against the above stated tremendous demands for effective policies to mitigate the multiple crises in existing global primary production systems, the success of SCs seems to be severely limited. In a nutshell, this is how the essence of this meta-study can be summarized against the background to the now broad societal debate about the benefits and shortcomings of SCs in governing bio-based global value chains.

As usual, limitations in terms of data collection and analysis also apply to this study. Systematic reviews generally need to strike a balance between the number of studies to include and the ability to categorize them in a meaningful way. Since the literature database, Scopus is specialized in the collection of scientific journal articles we chose it to select the literature for this this meta-study. We are not aware of any reasons how the use of Scopus could lead to a systematic bias regarding the inclusion of journal articles into our study. Nevertheless, there exists knowledge about the effectiveness of VSS that is not published in scientific outlets. The question is how much the validity of our study is limited by the fact that we excluded the so-called gray literature from the analyses. To control potential literature selection biases, we conducted an additional literature review using the online library, Evidensia, which specializes in systematically capturing the state of knowledge on the topic of SCs, including the gray literature. This review showed that the gray literature can be divided into three different groups: Academic literature not published in journals, reports by non-governmental organizations (NGOs) and other independent institutions, and reports drafted or commissioned by the SCs themselves (see supplementary material b).

As for the academic literature, we found ten additional pieces in the Evidensia database, most of which have preliminary characteristics (working papers, papers presented at conferences, research reports). Since researchers usually develop their preliminary findings into peer-reviewed articles, it is highly likely that much of the knowledge presented in these workings is also part of the corpus of peer-reviewed academic literature that we covered through our Scopus-based search. Further, we analyzed how the identified additional academic literature evaluated the overall success of SCs on the ground. Since the results (4 papers reached a skeptical conclusion, 3 papers reached mixed conclusions, 3 papers reached a favorable conclusion) largely match the results of our meta-study, we do not expect that the exclusion of the additional academic literature led to a significant bias in our analysis.

Another interesting type of literature that we did not include in our review are reports from independent third parties, such as international organizations, NGOs, or development agencies. In the Evidensia database, we found 22 reports from independent third parties that met our selection criteria. We excluded this literature from our meta-study because these third parties, even when working independently of SCs, typically either directly support or at least favor certain policy outcomes, which could influence their reporting. Nonetheless, we also coded how this literature evaluates the overall performance of SCs in practice. Again, the results (8 reports reached a skeptical conclusion, 9 reports reached mixed conclusions, 5 reports reached a favorable conclusion) largely confirm the picture in our meta-study, as only a few studies ended with a positive overall assessment, suggesting that also excluding this literature did not introduce significant bias into our meta-study.

Finally, we found 36 reports on the performance of SCs that were either directly drafted or commissioned by the SCs themselves. We analyzed a sub-sample of 10 reports and found an unusual, high amount of favorable assessments (6 out of 10 reports), with the remaining four cases ending with mixed results. We conclude from this analysis that due to severe conflicts of interest this literature seems to be highly biased towards positive assessments and should, therefore, be systematically omitted in a meta-study.

While we worked at length to develop and define a clear, objective coding scheme and analytical framework, paper coding remains subject to potential bias and human error. Categorizing evaluations of studies as favorable, mixed or skeptical is a strong simplification of often complex and differentiated scientific outcomes. However, given the urgent need to better understand the function of SCs as a possible pathway towards urgently-needed sustainability transformations, we accept this simplification tradeoff for the sake of summarizing the existing evidence in meaningful categories.

Another limitation refers to the differentiation of single studies into multiple cases when more than one country, dimension, certificate etc. was included in the same study. Since in these cases evaluations were reported by the same authors as part of the same publication, they may have undue influence on one other. However, this treatment of single studies as multiple cases, and inclusion of publications studying multiple conditions and outcome dimensions of sustainability certificates, facilitated a more nuanced analysis. Furthermore, this assessment represents

[1] https://www.oecd.org/agriculture/topics/water-and-agriculture/, last accessed at the 10th of Mai 2022.

a cross-sectional study and is therefore not suitable to capture cause and effect. Significant differences might be the result of other cofounding variables e.g. other characteristics of the included studies, not considered in our analysis. In this regard, it would also be interesting for future research to compare the reports of NGOs and other stakeholders with the results published in peer-reviewed journals. Despite the mentioned limitations, this study represents a comprehensive analysis of the current state of knowledge about the success of different certificates and the potential impact of different secondary variables.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Benjamin Cashore for his helpful feedback on an earlier draft. We also gratefully acknowledge funding by the German Federal Ministry of Food and Agriculture (BMEL) (grant number 2219NR291) through the research project 'Transformation and Sustainability Governance in South American Bioeconomies', the German Federal Ministry of Education and Research through the project LANUSYNCON under the FONA program (01UU2002), and the Ministry for Culture and Science of the State of North Rhine-Westphalia through the project TRANSSUSTAIN (grant number 005-1503-0008).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolecon.2022.107546.

References

Auld, Graeme, 2014. Constructing Private Governance. The Rise and Evolution of Forest, Coffee, and Fisheries Certification. Yale University Press, New Haven. Online verfügbar unter. http://www.jstor.org/stable/10.2307/j.ctt1bh4czv.

Auld, Graeme, Balboa, Cristina, Bernstein, Steven, Cashore, Benjamin, 2009. The emergence of non-state market-driven (NSMD) global environmental governance: A cross-sectoral assessment. In: Delmas, Magali A., Young Hg, Oran R. (Eds.), Governance for the Environment. New Perspectives. Cambridge University Press, Cambridge, pp. 183–218.

Barrientos, Stephanie, Smith, Sally, 2007. Do workers benefit from ethical trade? Assessing codes of labour practice in global production systems. Third World Q. 28 (4), 713–729. https://doi.org/10.1080/01436590701336580.

Bartley, Tim, 2007. Institutional emergence in an era of globalization: the rise of transnational private regulation of labor and environmental conditions. Am. J. Sociol. 113 (2), 297–351. https://doi.org/10.1086/518871.

Bartley, Tim, 2022. Power and the practice of transnational private regulation. New Polit. Econ. 27 (2), 188–202. https://doi.org/10.1080/13563467.2021.1881471.

Bastos, Lima, Mairon, G., Persson, U. Martin, Meyfroidt, Patrick, 2019. Leakage and boosting effects in environmental governance: a framework for analysis. Environ. Res. Lett. 14 (10), 105006. https://doi.org/10.1088/1748-9326/ab4551.

Baylis, Kathy, Honey-Rosés, Jordi, Börner, Jan, Corbera, Esteve, Ezzine-de-Blas, Driss, Ferraro, Paul J., et al., 2016. Mainstreaming impact evaluation in nature conservation. Conserv. Lett. 9 (1), 58–64. https://doi.org/10.1111/conl.12180.

Bennett, E.A., 2022. The efficacy of voluntary standards, sustainability certifications, andethical labels. In: Marx, A., van Calster, G., Wouters, J., Otteburn, K., Lica, D. (Eds.), Research Handbook on Global Governance, Business and Human Rights. Edward Elgar Publishing Limited, Cheltenham, UK, Northampton, Massachusetts, pp. 176–204.

Blackman, Allen, Rivera, Jorge, 2011. Producer-level benefits of sustainability certification. Conserv. Biol. 25 (6), 1176–1185. https://doi.org/10.1111/j.1523-1739.2011.01774.x.

Bloomfield, Michael John, 2012. Is Forest certification a hegemonic force? The FSC and its challengers. J. Environ. Dev. 21 (4), 391–413. https://doi.org/10.1177/1070496512449822.

Börner, Jan, Schulz, Dario, Wunder, Sven, Pfaff, Alexander, 2020. The effectiveness of Forest conservation policies and programs. Ann. Rev. Resour. Econ. 12 (1), 45–64. https://doi.org/10.1146/annurev-resource-110119-025703.

Bouslah, K., M'Zali, B., Turcotte, M.-F., Kooli, M., 2010. The impact of Forest certification on firm financial performance in Canada and the U.S. J. Bus. Ethics 96 (4), 551–572. https://doi.org/10.1007/s10551-010-0482-5.

Bowen, Glenn A., 2008. Naturalistic inquiry and the saturation concept: a research note. Qual. Res. 8 (1), 137–152. https://doi.org/10.1177/1468794107085301.

Boyatzis, Richard E., 2009. Transforming Qualitative Information. Thematic Analysis and Code Development. Sage Publications, Thousand Oaks (Ca.).

Bray, Joshua G., Neilson, Jeffrey, 2017. Reviewing the impacts of coffee certification programmes on smallholder livelihoods. Int. J. Biodiv. Sci. Ecosyst. Serv. Manag. 13 (1), 216–232. https://doi.org/10.1080/21513732.2017.1316520.

Carlson, K.M., Heilmayr, R., Gibbs, H.K., Noojipady, P., Burns, D.N., Morton, D.C., et al., 2018. Effect of oil palm sustainability certification on deforestation and fire in Indonesia. Proc. Nat. Acad. Sci. U.S.A. 115 (1), 121–126. https://doi.org/10.1073/pnas.1704728114.

Carlson, Anna, Palmer, Charles, 2016. A qualitative meta-synthesis of the benefits of eco-labeling in developing countries. Ecol. Econ. 127, 129–145. https://doi.org/10.1016/j.ecolecon.2016.03.020.

Cashore, Benjamin William, Newsom, Deanna, Auld, Graeme, 2004. Governing through Markets. Forest Certification and the Emergence of Non-state Authority. Yale University Press, New Haven. Online verfügbar unter. http://www.jstor.org/stable/10.2307/j.ctt1npqtr.

Cattau, M.E., Marlier, M.E., DeFries, R., 2016. Effectiveness of roundtable on sustainable palm oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015. Environ. Res. Lett. 11 (10) https://doi.org/10.1088/1748-9326/11/10/105007.

Corbin, Juliet M., Strauss, Anselm L., 2015. Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory, 4th edition. Sage, Los Angeles.

Dammert, Ana C., Mohan, Sarah, 2015. A survey of the economics of fair trade. J. Econ. Surv. 29 (5), 855–868. https://doi.org/10.1111/joes.12091.

DeFries, Ruth S., Fanzo, Jessica, Mondal, Pinki, Remans, Roseline, Wood, Stephen A., 2017. Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence. Environ. Res. Lett. 12 (3), 33001. https://doi.org/10.1088/1748-9326/aa625e.

Dietz, Thomas, Auffenberg, Jennie, Andrea, Estrella Chong, Grabs, Janina, Kilian, Bernard, 2018. The voluntary coffee standard index (VOCSI). Developing a composite index to assess and compare the strength of mainstream voluntary sustainability standards in the global coffee industry. Ecol. Econ. 150, 72–87. https://doi.org/10.1016/j.ecolecon.2018.03.026.

Dietz, T., Grabs, J., 2021. Additionality and implementation gaps in voluntary sustainability standards. New Polit. Econ. 27 (2), 1–22. https://doi.org/10.1080/13563467.2021.1881473.

Dietz, Thomas, Grabs, Janina, Chong, Andrea Estrella, 2019. Mainstreamed voluntary sustainability standards and their effectiveness: evidence from the Honduran coffee sector. Regul. Govern. https://doi.org/10.1111/rego.12239.

Dietz, Thomas, Andrea, Estrella Chong, Grabs, Janina, Kilian, Bernard, 2020. How effective is multiple certification in improving the economic conditions of smallholder farmers? Evidence from an impact evaluation in Colombia's Coffee Belt. J. Dev. Stud. 56 (6), 1141–1160. https://doi.org/10.1080/00220388.2019.1632433.

Ebbert, Daniel, 2019. Chisq.Posthoc.Test: A Post Hoc Analysis for Pearson's chi-Squared Test for Count Data. Online verfügbar Unter. https://CRAN.R-project.org/package=chisq.posthoc.test.

Eberlein, Burkard, Abbott, Kenneth W., Black, Julia, Meidinger, Errol, Wood, Stepan, 2014. Transnational business governance interactions: conceptualization and framework for analysis. Regul. Gov. 8 (1), 1–21. https://doi.org/10.1111/rego.12030.

Egels-Zandén, Niklas, Wahlqvist, Evelina, 2007. Post-partnership strategies for defining corporate responsibility: the business social compliance initiative. J. Bus. Ethics 70 (2), 175–189. https://doi.org/10.1007/s10551-006-9104-7.

Fortin, Elizabeth, Richardson, Ben, 2013. Certification schemes and the governance of land: enforcing standards or enabling scrutiny? Globalizations 10 (1), 141–159. https://doi.org/10.1080/14747731.2013.760910.

Fransen, L., 2012. Multi-stakeholder governance and voluntary programme interactions: legitimation politics in the institutional design of corporate social responsibility. Socioecon Rev. 10 (1), 163–192. https://doi.org/10.1093/ser/mwr029.

Fransen, Luc W., Kolk, Ans, 2007. Global rule-setting for business: a critical analysis of multi-stakeholder standards. Organization 14 (5), 667–684. https://doi.org/10.1177/1350508407080305.

Froese, Rainer, Proelss, Alexander, 2012. Evaluation and legal assessment of certified seafood. Mar. Policy 36 (6), 1284–1289. https://doi.org/10.1016/j.marpol.2012.03.017.

Garrett, Rachael D., Levy, Samuel, Gollnow, Florian, Hodel, Leonie, Rueda, Ximena, 2021. Have food supply chain policies improved forest conservation and rural livelihoods? A systematic review. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/abe0ed.

Grabs, Janina, 2020. Selling Sustainability Short? The Private Governance of Labor and the Environment in the Coffee Sector. Cambridge University Press (Organizations and the natural environment), Cambridge.

Hatanaka, Maki, Konefal, Jason, Constance, Douglas H., 2012. A tripartite standards regime analysis of the contested development of a sustainable agriculture standard. Agric. Hum. Values 29 (1), 65–78. https://doi.org/10.1007/s10460-011-9329-7.

Kassambara, Alboukadel, 2020. Ggpubr: ggplot2 Based Publication Ready Plots. Online verfügbar unter. https://rpkgs.datanovia.com/ggpubr/.

Llach, Josep, Marimon, Frederic, Alonso-Almeida, Del, Mar, María, 2015. Social accountability 8000 standard certification: analysis of worldwide diffusion. J. Clean. Prod. 93, 288–298. https://doi.org/10.1016/j.jclepro.2015.01.044.

de Córdoba Santiago Hg, Fernández, 2021. In: Marx, Axel, Dietz, Thomas, Elamin, Niematallah E.A. (Eds.), Better Trade for Sustainable Development. The Role of Voluntary Sustainability Standards. Geneva: United Nations.

Meemken, Eva-Marie, 2020. Do smallholder farmers benefit from sustainability standards? A systematic review and meta-analysis. Global Food Secur. 26, 100373. https://doi.org/10.1016/j.gfs.2020.100373.

Meemken, Eva-Marie, Qaim, Matin, 2018. Organic agriculture, food security, and the environment. Ann. Rev. Resour. Econ. 10 (1), 39–63. https://doi.org/10.1146/annurev-resource-100517-023252.

Meidinger, Errol, 2011. Forest certification and democracy. Eur. J. Forest Res. 130 (3), 407–419. https://doi.org/10.1007/s10342-010-0426-8.

Mena, Sébastien, Palazzo, Guido, 2012. Input and output legitimacy of multi-stakeholder initiatives. Bus. Ethics Q. 22 (3), 527–556. https://doi.org/10.5840/beq201222333.

O'Rourke, Dara, 2006. Multi-stakeholder regulation: privatizing or socializing global labor standards? World Dev. 34 (5), 899–918. https://doi.org/10.1016/j.worlddev.2005.04.020.

Oswaldo, Santos Baquero, 2019. Ggsn: North Symbols and Scale Bars for Maps Created with ggplot2 or Ggmap. Online verfügbar unter. https://github.com/oswaldosantos/ggsn.

Overdevest, Christine, Zeitlin, Jonathan, 2014. Assembling an experimentalist regime: transnational governance interactions in the forest sector. Regul. Gov. 8 (1), 22–48. https://doi.org/10.1111/j.1748-5991.2012.01133.x.

Oya, Carlos, Schaefer, Florian, Skalidou, Dafni, 2018. The effectiveness of agricultural certification in developing countries: a systematic review. World Dev. 112, 282–312. https://doi.org/10.1016/j.worlddev.2018.08.001.

Parkes, Graeme, Young, James A., Walmsley, Suzannah F., Abel, Rigmor, Harman, Jon, Horvat, Peter, et al., 2010. Behind the signs—a global review of fish sustainability information schemes. Rev. Fish. Sci. 18 (4), 344–356. https://doi.org/10.1080/10641262.2010.516374.

Pattberg, Philipp H., 2005. The Forest stewardship council: risk and potential of private forest governance. J. Environ. Dev. 14 (3), 356–374. https://doi.org/10.1177/1070496505280062.

Pendrill, Florence, Persson, U. Martin, Godar, Javier, Kastner, Thomas, Moran, Daniel, Schmidt, Sarah, Wood, Richard, 2019. Agricultural and forestry trade drives large share of tropical deforestation emissions. Glob. Environ. Chang. 56, 1–10. https://doi.org/10.1016/j.gloenvcha.2019.03.002.

Ponte, Stefano, 2012a. The marine stewardship council (MSC) and the making of a market for 'sustainable fish'. J. Agrar. Chang. 12 (2–3), 300–315. https://doi.org/10.1111/j.1471-0366.2011.00345.x.

Ponte, Stefano, 2012b. The marine stewardship council (MSC) and the making of a market for 'sustainable fish'. J. Agrar. Chang. 12 (2–3), 300–315. https://doi.org/10.1111/j.1471-0366.2011.00345.x.

Ponte, Stefano, 2014. 'Roundtabling' sustainability: lessons from the biofuel industry. Geoforum 54, 261–271. https://doi.org/10.1016/j.geoforum.2013.07.008.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. Vienna, Austria. Online verfügbar unter. https://www.R-project.org/.

Rasche, Andreas, 2012. Global policies and local practice: loose and tight couplings in multi-stakeholder initiatives. Bus. Ethics Q. 22 (4), 679–708. https://doi.org/10.5840/beq201222444.

Raynolds, Laura T., Murray, Douglas, Heller, Andrew, 2007. Regulating sustainability in the coffee sector: a comparative analysis of third-party environmental and social certification initiatives. Agric. Hum. Values 24 (2), 147–163. https://doi.org/10.1007/s10460-006-9047-8.

Reinecke, Juliane, Manning, Stephan, von Hagen, Oliver, 2012. The emergence of a standards market: multiplicity of sustainability standards in the global coffee industry. Organ. Stud. 33 (5–6), 791–814. https://doi.org/10.1177/0170840612443629.

Sachs, Jeffrey, 2015. The Age of Sustainable Development. Columbia University Press, New York. Online verfügbar unter. http://gbv.eblib.com/patron/FullRecord.aspx?p=1922296.

Schleifer, Philip, Sun, Yixian, 2020. Reviewing the impact of sustainability certification on food security in developing countries. In: Global Food Security, 24, p. 100337. https://doi.org/10.1016/j.gfs.2019.100337.

Sellare, Jorge, Meemken, Eva-Marie, Kouamé, Christophe, Qaim, Matin, 2020. Do sustainability standards benefit smallholder farmers also when accounting for cooperative effects? Evidence from Côte d'Ivoire. Am. J. Agric. Econ. 102 (2), 681–695. https://doi.org/10.1002/ajae.12015.

Seufert, Verena, Ramankutty, Navin, Foley, Jonathan A., 2012. Comparing the yields of organic and conventional agriculture. Nature 485 (7397), 229–232. https://doi.org/10.1038/nature11069.

Terstappen, Vincent, Hanson, Lori, McLaughlin, Darrell, 2013. Gender, health, labor, and inequities: a review of the fair and alternative trade literature. Agric. Hum. Values 30 (1), 21–39. https://doi.org/10.1007/s10460-012-9377-7.

Tröster, Rasmus, Hiete, Michael, 2018. Success of voluntary sustainability certification schemes – a comprehensive review. J. Clean. Prod. 196, 1034–1043. https://doi.org/10.1016/j.jclepro.2018.05.240.

Vanderhaegen, Koen, Akoyi, Kevin Teopista, Dekoninck, Wouter, Jocqué, Rudy, Muys, Bart, Verbist, Bruno, Maertens, Miet, 2018. Do private coffee standards 'walk the talk' in improving socio-economic and environmental sustainability? Glob. Environ. Chang. 51, 1–9. https://doi.org/10.1016/j.gloenvcha.2018.04.014.

Vogel, David, 2008. Private global business regulation. Annu. Rev. Polit. Sci. 11 (1), 261–282. https://doi.org/10.1146/annurev.polisci.11.053106.141706.

Weber, J.G., 2011. How much more do growers receive for Fair Trade-organic coffee? Food Policy 36 (5), 678–685. https://doi.org/10.1016/j.foodpol.2011.05.007.

Wickham, Hadley, 2016. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York. Online verfügbar unter. https://ggplot2.tidyverse.org.

Wickham, Hadley, 2019. Tidyverse: Easily Install and Load the Tidyverse. Online verfügbar unter. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, 2021. Forcats: Tools for Working with Categorical Variables (Factors). Online verfügbar unter. https://CRAN.R-project.org/package=forcats.

Wijen, Frank, 2014. Means versus ends in opaque institutional fields: trading off compliance and achievement in sustainability standard adoption. AMR 39 (3), 302–323. https://doi.org/10.5465/amr.2012.0218.

Wilke, Claus O., 2020. Cowplot: Streamlined Plot Theme and Plot Annotations for ggplot2. Online verfügbar unter. https://wilkelab.org/cowplot/.